Examining the Relationship Between Success For All and Non-Cognitive Outcomes

by

Elizabeth Kim

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

April, 2018

Abstract

The current education policy and research context recognizes non-cognitive factors as important educational outcomes to consider in addition to academic achievement outcomes. This study examines whether a widely used whole-school reform program, Success For All (SFA), is linked to greater student non-cognitive outcomes, namely engagement, self-efficacy, and antisocial behavior, compared to a comparison group. The study further seeks to clarify the relationship between achievement and non-cognitive factors over time. Using multilevel models with propensity scores and autoregressive cross-lagged panel models to examine possible mediation with a diverse sample of SFA (*n* = 469) and control (*n* = 508) students, this study finds a small but significant positive effect of SFA on teacher-reported student engagement but no meditating relationships between non-cognitive factors and reading achievement. The results of the study suggest the promise of programs that are not explicitly SEL-focused in improving students' non-cognitive outcomes and engagement in particular. Implications for practice and directions for future research are offered.

Primary Reader: Robert E. Slavin

Secondary Readers: Lieny Jeon, Marcia Davis

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Background

What does it mean to be an educated person? Traditionally, and particularly during the No Child Left Behind era, to be educated has been seen as synonymous with high achievement test scores (Duckworth, Quinn, & Tsukayama, 2012; Noddings, 2005). In contrast, the current educational zeitgeist has emphasized the idea of educating the "whole child" (Humphrey, 2013; Noddings, 2005; Zins, Weissberg, Wang, & Walberg, 2004), paying special attention to not only the cognitive but also the non-cognitive aspects of student learning. While there is no shortage of definitions and conceptualizations of what non-cognitive factors mean, it is generally represented as "sets of behaviors, skills, attitudes, and strategies that are crucial to academic performance in [students'] classes" (Farrington et al., 2012). Others have used terms such as social and emotional learning, 21st century skills, and soft skills to refer to the same conceptual space (Duckworth & Yeager, 2015).

Recent moves in education policy and research have brought the importance of non-cognitive factors to the forefront. For instance, the Every Student Succeeds Act (ESSA) now requires at least one nonacademic indicator in its school evaluation measures. Individual states have also incorporated non-cognitive factors in states' educational standards: the state of Illinois has adopted comprehensive K-12 social and emotional learning standards, the District of Columbia Public Schools lists educating the whole child as part of their district-wide strategic plan, and states such as Kansas, Tennessee, Vermont, and Washington address specific parts of SEL in their standards documents for subjects such as communication and service learning.

Research in the non-cognitive space has also proliferated in the last few decades (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Federal Investing in Innovation (i3) grants, which develop and evaluate innovations in education, have included studies on mindfulness and the effectiveness of specific social and emotional programs. Research on characteristics such as grit and self-control (Duckworth, Peterson, Matthews, & Kelly, 2007), character (Berkowitz & Bier, 2007), mindfulness (Schonert-Reichl & Lawlor, 2010), and social and emotional learning (SEL; Zins, Weissberg, Wang, & Walberg, 2004) in general have galvanized public attention. It is not surprising that the discussion of non-cognitive factors has gained much traction in recent years given the robust literature that is resurfacing and the current research literature that assert the importance of non-cognitive factors in their own right (Garcia, 2016) as well as in their role of predicting various academic outcomes, labor productivity and earnings, and other well-being measures (Durlak et al., 2011; Farrington et al., 2012; Heckman & Rubinstein, 2001; Murnane, Willet, Braatz, & Duhaldeborde, 2001; Payton et al., 2008).

To date, there is no definitive list of non-cognitive factors that are thought to be important in education. The current study draws on a widely cited framework of non-cognitive factors (Farrington et al., 2012) that groups various non-cognitive factors into five interrelated categories. This study focuses on student engagement, self-efficacy, and antisocial behavior as non-cognitive factors of particular importance in education. Student engagement is commonly understood as involvement in and with learning activities (Fredricks, Blumenfeld, & Paris, 2004). Antisocial behavior refers to both physical and non-physical disruptive acts (Rosen, Glennie, Dalton, Lennon, & Bozick, 2010). Self-efficacy (Bandura, 1986) refers to a person’s psycho-social beliefs in his or her ability to succeed in a task or subject. These variables were selected among many different non-cognitive factors because they are not only important goals in their own right but also because a growing body of research indicates their relationship with important school outcomes such as better academic achievement, school retention, and lower risk of delinquency (Gutman & Schoon, 2013; Klem & Connell, 2004; Multon, Brown, & Lent, 1991) and because they represent a range of non-cognitive categories of interest.

Statement of The Problem

As schools face increasing pressure to improve academic performance as well as to be attentive to children's non-cognitive needs, it is imperative that educators find and implement effective approaches that optimize growth, especially in non-cognitive areas. A need for rigorous evidence of programs that can promote non-cognitive growth is particularly apt for the current policy context, which emphasizes both the need to assess non-academic indicators and the need for evidence-based programs.

There is a growing body of research suggesting the effectiveness of whole school reform models in promoting students' non-cognitive factors. Whole school reforms move beyond making isolated programmatic changes to involving coordinated changes in instructional management, practices, and relationships among stakeholders as a way to improve student outcomes (Corsello & Sharma, 2015; Rowan, Correnti, Miller, & Camburn, 2009). Among the most widely used whole school approaches in the U.S. is Success For All. Success For All (SFA) highlights three general elements: 1) research-based, prescribed curriculum in the areas of reading, writing, and language arts, 2) whole school improvement components that address non-instructional issues, and 3) strategies to secure teacher buy-in, provide school personnel with initial and ongoing training, and foster shared school leadership (Quint et al., 2013).

Although widely known for its strong curricular features in reading, SFA also entails whole school improvement components thought to influence non-cognitive growth such as efforts to increase family involvement and community engagement, facilitating students' social skills school-wide through a “Getting Along Together” social and emotional skills program, and increased staff support for behavior (Quint et al., 2013). Based on other whole school reforms that showcase similar features (e.g. School Development Program, Positive Action, School-wide Positive Behavior System) described in Chapter 2, there is good reason to believe that such a whole school improvement approach may not only impact students’ academic but also non-cognitive development.

A great deal of research has documented Success For All’s effectiveness in improving students’ academic achievement (e.g., Borman et al., 2003; Chamberlain, Daniels, Madden & Slavin, 2007; Quint, Zhu, Balu, Rappaport, & DeLaurentis, 2015). Yet there is a gap in the literature as to whether Success For All also has an impact on students’ non-cognitive factors, such as engagement, antisocial behavior, and self-efficacy. The current context of evidence-based policy and increased emphasis on non-academic measures warrant a rigorous examination of Success For All’s impact on non-cognitive outcomes.

An additional gap in the literature is that the directional relationship between non-cognitive factors and academic achievement is empirically unclear in the literature. While much of the research has assumed that non-cognitive factors impact achievement over time, the reverse direction may also be true. Many scholars have acknowledged that there is likely a bidirectional relationship between the two outcomes (Cunha & Heckman, 2007; Farrington et al. 2012), but there is little empirical evidence that tests such relationships.

Relatedly, the mechanisms by which school based programs influence non-cognitive factors are not well understood. That is, how do programs impact student non-cognitive outcomes? This study proposes one pathway through which Success For All may influence the non-cognitive domain: academic achievement. While there is an abundance of research on

Success For All's impact on academic achievement as a separate outcome, none to date have investigated whether these outcomes are further linked to non-cognitive factors. There is good reason to believe that academic achievement may act as a mediator of program effects on non-cognitive factors based on the literature that links achievement to non-cognitive development (e.g. Cook, Murphy, & Hunt, 2000; Eliot, Cornell, Gregory, & Fan, 2010; Jimerson & Ferguson, 2007). Understanding the mechanism through which program effects are achieved may shed light on the process by which schools can shape non-cognitive outcomes.

The Current Study

Considering the current policy context and gaps identified in the literature, this study examines the impact of the Success For All program (SFA; Slavin, Madden, Chambers, & Haxby (2009), 2009) on children's non-cognitive development and the mechanisms through which it may achieve this impact. It also seeks to clarify the relationship between non-cognitive factors and achievement over time. The current study seeks to answer the following questions with subsequent hypotheses:

1. Are there differences in non-cognitive factors among students in SFA and non-SFA schools over time?

   H1: It is hypothesized that children in SFA will score higher on non-cognitive factors.

2. To what extent does participation in SFA relate to growth in non-cognitive factors over time?

   H2: It is hypothesized that students participating in SFA will have higher rates of growth in non-cognitive factors compared to non-SFA students.

3. To what extent does achievement predict non-cognitive outcomes over time, regardless of SFA status?

   H3: It is hypothesized that achievement will positively predict non-cognitive outcomes.

4. To what extent do non-cognitive factors predict achievement over time, regardless of SFA status?

   H4: It is hypothesized that non-cognitive factors will positively predict achievement.

5. To what extent is the effect of SFA on students' non-cognitive factors mediated by academic achievement?

   H5: It is hypothesized that SFA schools will have higher ratings of academic achievement, which in turn will positively impact students' non-cognitive factors.

Definition of Terms

- Engagement: students' behavioral, emotional, and cognitive involvement in and with their learning activities (Fredricks, Blumenfeld, & Paris, 2004).
- Antisocial behavior: physical and non-physical disruptive acts (Dalton, 2010).
- Self-efficacy: psycho-social beliefs a student has in his or her ability to succeed in a task or subject (Bandura, 1986).
- Whole school reform: approaches to improving outcomes in entire schools with the school as the primary unit of intervention (Slavin, 2008).
- Academic achievement: performance outcomes that indicate the extent to which a person has accomplished specific goals in the instructional context (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014).

Procedures

To answer the research questions, data come from the Study of Instructional Improvement (SII), conducted by the University of Michigan and the Consortium for Policy Research in Education, which evaluated the effectiveness of three whole school reform models (America's Choice, Accelerated Schools, and Success For All) compared to control schools in improving achievement using a longitudinal, quasi-experimental design. The current study uses only the comparison between Success For All schools and control schools in the original sample as the other two school reform models are no longer in use. Each school was followed for three years from kindergarten to the end of second grade, and teacher, student, parent, and school leader surveys were administered each spring. Students were also assessed in reading and math at the beginning and end of each year.

The outcome measures were derived from teacher and student reports of non-cognitive factors. Students reported their engagement and self-efficacy, and teachers reported on individual students' engagement and disruptive behavior in class. Covariates were derived from school records of student characteristics as well as school-level characteristics.

Because schools were not randomly assigned to treatment, propensity scores were used to adjust for selection bias and strengthen arguments of causal effects associated with the program. Propensity scores control for selection into treatment based on observable variables for a more reliable estimate of changes in non-cognitive factors by treatment. Additionally, hierarchical linear models (HLM) were used to answer research questions 1 and 2. Such techniques are more appropriate than traditional linear regression when the data are nested, since they estimate more accurate standard errors and allow for partitioning between individual and school level variance (Raudenbush & Bryk, 2002). Auto-regressive cross-lagged panel models were used to answer research question 3 through 5. The longitudinal nature of the data allows examination of how different variables develop and impact other variables over time.

Significance of the Study

A holistic view of educational outcomes suggests that non-cognitive factors such as engagement, self-efficacy, and behavior regulation are important learning objectives for students. Accordingly, the current policy context urges the use of evidence-based programs that can demonstrate impact on non-cognitive as well as academic factors. While Success For All has shown strong evidence of effectiveness on student achievement, it is plausible that its whole school components may also influence students' non-cognitive growth. However, this possibility has not been tested. The current study proposes to conduct a rigorous examination of the program's impact on non-cognitive factors. In this way, this study will be of particular interest to practitioners and policymakers who are invested in finding effective programs that enhance students' non-cognitive factors. Additionally, clarifying the relationship between non-cognitive factors and achievement over time will contribute to the overall literature that seeks to add more empirical evidence to the two theoretically-linked concepts. Finally, this study seeks to extend the theoretical literature on *how* such program effects may come about. By exploring a pathway to non-cognitive outcomes – academic achievement – this study hopes to shed light on a mechanism through which changes in these outcomes occur. Achievement may not only be an important outcome of programs but also a mediator for non-cognitive outcomes. Thus, this study seeks to offer both important practical and theoretical implications.

Overview of the Study

Chapter 1 has presented the introduction, statement of the problem, research questions, definition of terms, and significance of the study. Chapter 2 contains the review of related literature and research related to the problem being investigated. In particular, it gives an overview of the importance of non-cognitive factors, different approaches to improving non-cognitive factors, a detailed description of the Success For All program, and gaps in the current literature that inform the research questions for this study. The methodology and procedures used to gather data for the study are presented in Chapter 3. The results of analyses and findings to emerge from the study will be outlined in Chapter 4. Finally, Chapter 5 will contain a summary of the study and findings, conclusions drawn from the findings, a discussion, and recommendations for further study.

Chapter 2. Literature Review

Part 1: Overview of Non-Cognitive Factors

Historical overview

Traditionally, test scores of cognitive ability have been prioritized over other measures in school (Duckworth, Quinn, & Tsukayama, 2012; Noddings, 2005). Reflecting on the tradition of scholars to emphasize the cognitive while neglecting other traits, Heckman and Rubinstein (2001) write:

> The early literature on human capital (e.g. Becker, 1964) contrasted cognitive-ability models of earnings with human capital models, ignoring non-cognitive traits entirely. The signaling literature (e.g., Spence, 1974) emphasized that education was a signal of a one-dimensional ability, usually interpreted as a cognitive skill... Widespread use of standardized achievement and ability tests for admissions and educational evaluation are premised on the belief that the skills that can be tested are essential for success in schooling, a central premise of the educational-testing movement since its inception. (p. 145)

In contrast to this traditionally narrow assumption of what is important for schooling, common sense suggests that other factors such as personality traits, persistence, motivation, and the ability to get along with others also matter for success in school and in life. A century ago, Binet and Simon (1916) noted that performance in school “admits other things than intelligence; to succeed in his studies, one must have qualities which depend on attention, will, and character” (p. 254). Bowles and Gintis (1976) were among the first to popularize the phrase “non-cognitive traits," marking the distinction between such traits and strictly academic skills. Coming from a predominantly economic standpoint, they suggested non-academic traits such as agreeableness, motivation, and helpfulness, in addition to academic skills, as determinants of labor market success. Thus, they defined non-cognitive traits as "employer-valued attributes" that also had parallels in the classroom. Since then, there has been a resurgence of interest in non-cognitive traits as more researchers, policymakers, and educators acknowledge the importance of traits that are not traditionally tested. More current definitions include the University of Chicago Consortium on School Research (CCSR) description of non-cognitive factors as "sets of behaviors, skills, attitudes, and strategies that are crucial to academic performance in [students'] classes" (Farrington et al., 2012, p. 2). The current study adopts this definition of non-cognitive factors.

Importance of non-cognitive factors

It is important to note in reviewing the literature on non-cognitive factors and student outcomes that there is a lack of clarity on the direction of causality. Studies examining non-cognitive factors of children and adolescents are predominantly correlational, since it is difficult to manipulate such variables in an experimental setting (Rosen et al., 2010; Gutman & Schoon, 2013). Moreover, the quality of research that allows for claims of causality varies. Many studies present a simple correlation between non-cognitive factors and outcomes such as academic achievement, without controlling for pretests or other confounding factors that may explain the relationship such as socioeconomic status or cognitive ability (Hinshaw, 1992; Morgan, Farkas, Tufis, & Sperling, 2008). Admittedly, the relationships are difficult to disentangle completely (Kirsch et al., 2003). However, in reviewing the literature linking non-cognitive factors and important student outcomes here, I highlight those studies which have attempted to control for potential confounding variables to get closer to claims of causality.

Students who possess effective learning strategies and positive academic mindsets such as a sense of belonging, self-efficacy, and motivation are likely to succeed in their school courses because of increased persistence and positive academic behaviors (Farrington et al., 2012). Moreover, other non-cognitive factors such as organizational skills and the ability to work well with others are valued by teachers and may cause them to more highly rate students (Lee & Shute, 2009). For instance, using ECLS-K data Cornwell, Mustard, and Parys (2013) found that teachers gave better grades to students who displayed more positive approaches to learning even when their standard achievement scores were comparable. That is, students with more positive non-cognitive skills are likely to succeed in school because they have the mindsets and social skills that are valued by educators and are important for academic achievement. Empirical studies that affirm the association between non-cognitive factors and achievement abound. For instance, Durlak et al. (2011) conducted a meta-analysis of over 200 interventions aimed at increasing the social and emotional learning (SEL) of children from kindergarten through high school (ages 5–18). Limiting the interventions to experimental studies with control groups, Durak et al. (2011) found that on average, students in SEL programs had higher academic achievement than control students. The estimated gain in academic performance was 11 percentile points. However, as noted previously, these figures should be taken with caution since the review added little detail as to what was controlled for (e.g. pretest achievement levels, race, SES), especially among non-randomized studies.

Beyond academics, the literature has documented the positive relationship between non-cognitive factors and more distal outcomes such as economic success (e.g. Heckman &

Rubinstein, 2001; Heckman, Stixrud, & Urzua, 2006) and civic engagement (e.g. Hillygus, Holbein, & Snell, 2016; Gutman & Schoon, 2013). Sociologists and economists have particularly been interested in non-cognitive behaviors of workers and their association with increased labor-market rewards (Lee & Shute, 2009). This is because non-cognitive traits such as the ability to work with others and organizational skills are highly valued among employers (Lee & Shute, 2009). To illustrate, Heckman and Rubinstein (2001) argued that GED holders earn less than high-school graduates because the GED presents a "mixed signal" to employers. They argued that GED holders were equivalent to high school graduates in cognitive skills (measured by the Armed Forces Qualifying Test) but lacked in non-cognitive factors such as the ability to think ahead, to persist in tasks, or to adapt to their environments. Heckman and Rubinstein (2001) observed that holding cognitive ability constant, GED holders earned less than high school graduates in the long run because the GED signaled to employers that the students lacked these important non-cognitive skills. Thus, the authors attributed differences in earnings to differences in factors outside of cognitive ability.

Clearly, the research underlines the need to pay attention to non-cognitive factors as important outcomes of schooling in addition to traditional academic measures.

Defining non-cognitive factors

Some lament the term "non-cognitive factors" because it suggests that such skills and attitudes preclude any sort of cognitive work, while it is known that very few human behaviors are completely divorced from cognition (Borghans, Duckworth, Heckman, & Weel, 2008). Others have suggested terms such as social and emotional learning, 21st century skills, and soft skills to refer to the same conceptual space (Duckworth & Yeager, 2015). Despite the Tower of Babel the term has come to represent in some way, Duckworth and Yeager (2015) note common characteristics shared among the different terms, in particular that they are: "(a) conceptually independent from cognitive ability, (b) generally accepted as beneficial to the student and to others in society, (c) relatively rank-order stable over time in the absence of exogenous forces (e.g., intentional intervention, life events, changes in social roles), (d) potentially responsive to intervention, and (e) dependent on situational factors for their expression" (p. 239). I use the term "non-cognitive factors" here as an umbrella term that encapsulates the various behaviors, skills, attitudes, and strategies crucial to academic success.

In addition to various terminology used to describe non-cognitive factors, there are differing opinions as to which skills and attitudes should count as non-cognitive factors. Currently, there is no definitive list of non-cognitive factors that are relevant to the education process (Garcia, 2016). For instance, the Collaborative for Academic, Social, and Emotional Learning (CASEL) describes socio-emotional learning in terms of 5 competencies (self-awareness, self-management, social awareness, relationship skills, and responsible decision making) while the National Research Council categorizes 21st century skills into cognitive, intrapersonal, and interpersonal skills. Meanwhile, the Education Policy Institute put forth a list of non-cognitive skills that included traits such as emotional health, social skills, work ethic, and community responsibility (Garcia, 2016).

The CCSR model of non-cognitive factors

Among the various conceptualizations of necessary skills and attitudes, this study adopts one model of non-cognitive factors, University of Chicago Consortium on Chicago School Research (CCSR)'s model, for two main reasons: 1) it is a widely used and accepted model and, 2) it represents a great breadth of non-cognitive factors that is presented in a logically organized schema. The CCSR synthesizes the literature on non-cognitive factors into a model of five interrelated categories that influence student performance in school: academic behaviors, academic perseverance, academic mindsets, learning strategies, and social skills (Farrington et al., 2012).

First, academic behaviors refer to behaviors such as regularly attending class, paying attention, and participating in instructional activities that are illustrative of being a "good student." Observable non-cognitive factors such as effort and attendance are part of this category.

Second, academic perseverance refers to students' tendencies to complete school assignments to the best of their ability. Non-cognitive factors such as grit or delayed gratification are examples in this category.

Third, academic mindsets represent attitudes or beliefs students have about themselves in relation to academic work. Non-cognitive factors such as self-efficacy, growth mindsets, and sense of belonging are included in this category.

Fourth, learning strategies include processes and tactics that facilitate cognitive work such as thinking, learning, or organizing. Non-cognitive factors such as metacognition, self-regulated learning, and goal setting represent this category.

Finally, social skills include interpersonal qualities that allow one to work well with others and manage one's own behavior in class. Non-cognitive factors such as empathy, pro-social behavior, and self-control are included in this category. The CCSR model assumes that many non-cognitive factors within these categories are mutually reinforcing and maintain reciprocal relationships. Although the framework hypothesizes different pathways through which these five categories may influence one another and academic achievement, the model is more useful in this study for its conceptual organization of different non-cognitive skills. The figure below illustrates the CCSR model.

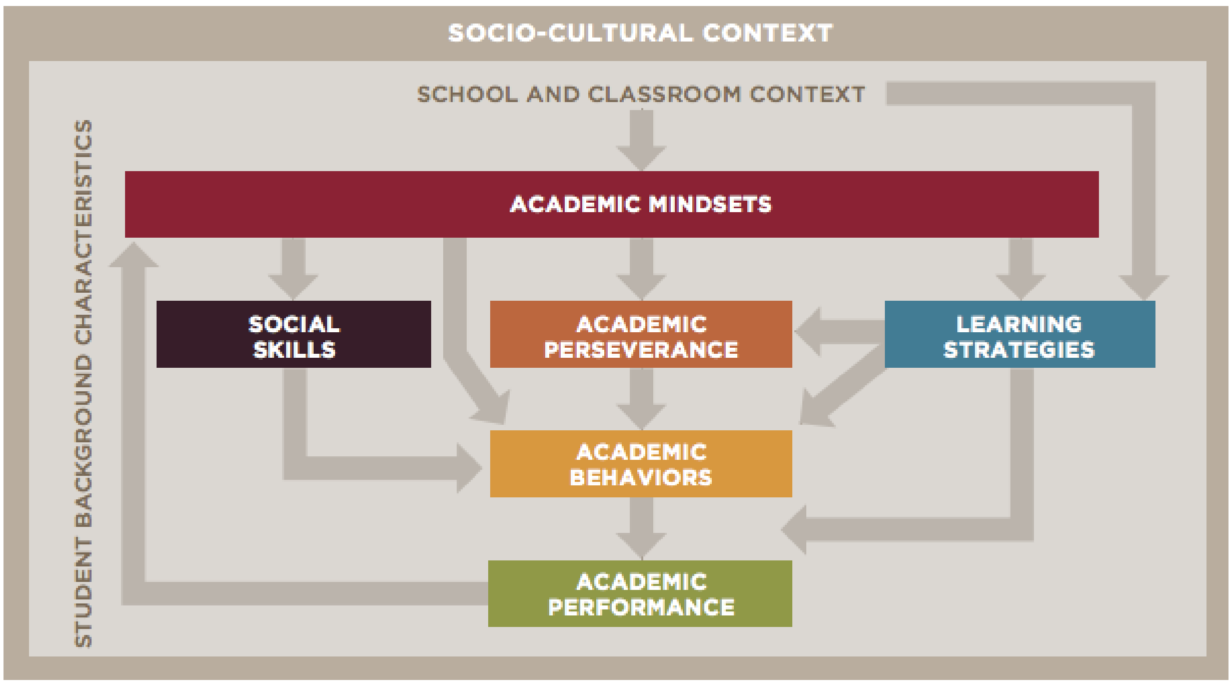

Figure 1. CCSR model of non-cognitive factors (Farrington et al., 2012, p. 12)

Non-cognitive factors in this study

Among the many types of non-cognitive factors that are suggested in the literature and the CCSR model, this study focuses on three factors as of particular importance in education: engagement, antisocial behavior, and self-efficacy. These factors were chosen for two primary reasons: 1) the research linking these factors with achievement is robust, and 2) the three selected represent a breadth of categories within the CCSR framework, highlighting the diversity of non-cognitive factors. The following section briefly reviews the definitions and research related to the three constructs.

1. Student engagement.

Definition. Student engagement is commonly understood as a meta-construct consisting of students' behavioral, emotional, and cognitive involvement in and with their learning activities (Fredricks, Blumenfeld, & Paris, 2004). Behavioral engagement draws on the idea of participation, on-task behavior, and completion of work; emotional engagement involves positive and negative reactions to teachers, classmates, academics, and/or school; cognitive engagement represents thoughtfulness and effort in comprehending complex ideas and skills. Student engagement is thought to be part of the academic behaviors category of the CCSR framework (Farrington et al., 2012). It is important to note that while related, motivation and engagement do not denote the same concept. Whereas motivation involves the psychological processes that may direct one's actions, engagement is described as "energy in action" that reflects a students' *active involvement* or participation in a task (Appleton, Christenson, & Furlong, 2008).

Related outcomes. Interest in engagement has burgeoned in the past decade because of the growing body of evidence that supports its relationship with desirable student outcomes and because it is thought to be a malleable trait (Fredricks et al. 2004).

Achievement. Engaged students are more likely to attain higher grades (Klem & Connell, 2004; National Research Council & Institute of Medicine, 2004; Wang & Pomerantz, 2009). Large-scale assessments such as the National Educational Longitudinal Study (NELS) have documented the positive link between engagement and achievement. For instance, Finn and Rock (1997) used data from the National Educational Longitudinal Study of 1988 (NELS: 88) to show a strong relationship between student engagement and academic achievement. A sample of 8 to 12th graders were divided into resilient or nonresilient groups based on their academic achievement and school completion rates, with resilient students representing the higher achieving, school completing group. After controlling for family structure, socioeconomic status, and prior school experiences the authors found significant differences between the resilient and nonresilient student groups in student- and teacher-rated measures of engagement indicators such as working hard, being prepared, attending school, and participating in extracurricular activities.

Domain specific studies have highlighted this link between engagement and achievement as well. Guthrie at al. (1999) found that third- and fifth-grade students' behavioral engagement, or self-reported time spent reading in and out of school, was positively associated with students' reading comprehension scores, even while controlling for potentially confounding variables such as background knowledge, previous grades, intrinsic motivation, and self-efficacy.

Similarly, in an experimental study, Wigfield et al. (2008) examined that the effects of CORI, an instructional approach to improve students' reading through motivation, on fourth-grade students' reading comprehension. They found that the effects of the approach on students' reading achievement were fully mediated by students' behavioral engagement in reading. That is, when reading engagement was controlled for, the differences between treatment and control students in achievement were no longer significant. This suggests that the instructional approach increased students' engagement, which fully accounted for the increase in reading achievement.

Longitudinally, Ladd and Dinella (2009) sought to determine how early engagement patterns may predict achievement. They found that students who reported higher behavioral and emotional engagement in grades 1 to 3 made greater academic gains by the 8th grade than those who displayed lower levels of engagement, statistically controlling for reading achievement in grade 1. Put another way, increasing behavioral engagement produced a positive slope for achievement, controlling for prior achievement. The authors concluded with the observation that engagement is an important prerequisite for achievement.

Dropout. Engagement is further theorized to relate to dropout. Finn's (1989) participation–identification model asserts that patterns of engagement and disengagement in the early grades will impact students' behavior and academic achievement in the later years. The model predicts that lack of participation in school activities (i.e., lack of behavioral engagement) leads to unsuccessful school outcomes, which further leads to emotional withdrawal and lack of identification with the school. The cycle continues such that lack of identification breeds lack of participation and lower school achievement, potentially leading to eventual dropout.

Indeed, there is some evidence that supports this theory. The Beginning School Study evidenced the longitudinal effects of early school behaviors on dropout rates (Alexander, Entwisle, & Horsey, 1997). In this study, teachers' ratings of children's behavioral engagement in the first grade were related to the decision to drop out of high school. Evidence from longitudinal data suggests that school engagement has been associated with the prevention of delinquency, school dropout, and substance use (Li et al., 2011; Wang & Fredricks, 2014).

Finn's (1989) model also assumes that students' emotional engagement has impact on the decision to drop out, in that alienation, or feelings of estrangement and social isolation, may influence their decision to drop out. Emotional engagement in school is associated with children's emotional adjustment, regardless of prior academic achievement (Gutman, Brown, Ackerman, & Obolenskaya, 2010). Ethnographic studies also lend support to this claim in that an emotional connection to the school or teachers can act as a protective factor, especially for at-risk students to remain in school (Fine, 1991; Mehan, Villanueva, Hubbard, Lintz, Okamato, &

Adams, 1996).

Thus, the literature on student engagement has both proximal and distal links to student outcomes such as achievement and dropping out. It is also presumed to be malleable and thus subject to change by the context and environment (Fredricks, Blumenfeld, & Paris, 2004).

2. Antisocial behavior.

Definition. Students can display disruptive behavior, which commonly consists of both physical acts (e.g., violence to others or to objects in the environment, disruptive or intentionally distracting actions) and nonphysical acts (e.g., exclusion, rejection, humiliation, any form of verbal abuse) (Dalton, 2010). Antisocial behavior is thought to be part of the social skills category of the CCSR framework (Farrington et al., 2012). It is conceptually distinct from personality disorders or conduct disorders that can be classified by the *Diagnostic and Statistical Manual of Mental Disorders* (Dalton, 2010).

Related outcomes. Studies examining antisocial or prosocial behaviors have linked antisocial behavior with psychological problems and lower academic achievement (Dalton, 2010; Miles & Stipek, 2006).

Achievement. Students with antisocial behavior are likely to exhibit lower academic achievement and rates of academic engaged time (Coie & Jacobs, 1993; Lane & Wehby, 2002). Wentzel (1993) used data from 423 students in 6th and 7th grade to find that prosocial and antisocial behavior were significantly related to GPA and standardized test scores. Multiple regression analyses revealed that both prosocial and antisocial behavior were independent predictors of GPA, controlling for a number of confounding variables such as IQ, academically oriented behavior, family structure, teachers' preference for students, and days absent from school.

Chen et al. (2010) used a longitudinal sample of students from grades 3 to 6 and structural equation modeling to find that aggression predicted later social competence and academic achievement after accounting for earlier levels of social competence and achievement. More specifically, Chen et al. (2010) found that aggression in grades 2 to 4 had significant direct effects on later social and academic outcomes, and that the effects of aggression were more pronounced in the younger grades. The longitudinal panel design used in the study allowed analysis of cross-lagged direct and indirect effects of aggression on social and academic variables at various time points.

Bierman et al. (2013) investigated factors that predict later achievement among 891 children with aggressive-disruptive behavior problems at school entry. They undertook a series of multilevel models that suggested that controlling for concurrent cognitive skills, the degree of initial aggression at kindergarten significantly predicted school difficulty such as low GPA in the elementary years and grade retention in the secondary years. In fact, kindergarten aggression was the only significant predictor of grade retention in secondary school. The authors concluded that initial aggression at school entry accounted uniquely for later school maladjustment.

Emotional and social outcomes. Students with antisocial behavior issues are also often alienated from friends and teachers, reducing overall social competence and posing a risk for the development of other emotional problems and delinquency (Schaeffer, Pteras, Jalongo, Poduska, & Kellam, 2003). For example, Stipek and Miles (2008) used growth curve modeling with a longitudinal sample of students to find that the child aggression predicted conflictual relationships with teachers, which mediated the relationship between aggression and negative academic outcomes. That is, student aggression was likely to elicit negative reactions from teachers and increase interpersonal conflict over time.

White and Loeber (2008) used a longitudinal sample of boys from the The Pittsburgh Youth Study (PYS) to examine the links between early aggression and social outcomes. They found that controlling for family background, neighborhood, and achievement, aggressive behavior predicted the student being teased or disliked by peers. In other words, students who were rated as aggressive by their teachers were more likely to be socially marginalized by peers. These negative emotional and relational outcomes are particularly problematic because they can lead to lower motivation and interest, and consequently worse academic performance (Farrington et al., 2012), offering one possible explanation for lower achievement among students with behavioral issues.

Such findings indicate that aggression and antisocial behavior can alter emotional and social conditions such that academic performance is compromised. Furthermore, the negative impacts of antisocial behavior are thought to extend beyond academics. Prosocial behavior, such as getting along with others and collaborative skills, are highly valued in the school but also in the workplace. Casner-Lotto and Barrington (2006) conducted a study that surveyed over 400 employers in the United States, who ranked the desired skill sets needed for new entrants to the workforce. The results revealed that the four most important skills for employers were oral communication, teamwork/collaboration, professionalism/work ethic, and critical thinking/problem solving. On the contrary, of a list of 20 skills, mathematics, science, and history/geography were ranked 15th, 16th, and 19th, respectively. The study highlights the importance of social skills such as teamwork and collaboration in the workplace, whereas antisocial behavior represents the antithesis of such skills.

Differences by group. It is important to note, however, that antisocial behavior is interpreted differently by various groups of students. For instance, aggression is sometimes associated with positive social outcomes such as popularity and self-esteem, especially among boys (Becker & Luthar, 2007). Although popularity and self-esteem are not always protective factors for academic achievement, studies do suggest antisocial behaviors can operate through changed social relationships, both in positive and negative ways that are dependent on the social group (e.g., boys versus girls) (Dalton, 2010).

In addition to differences between boys and girls, race has been often cited as a factor in perceived antisocial behavior. It is widely known that minority students experience a disproportionate amount of disciplinary action in school because of behavior issues (Gregory, Skiba, & Noguera, 2010). Data from the National Center for Education Statistics (NCES) indicates that 57 percent of high-school African American males were suspended in 2007, the highest number among any racial or gender group. The racial and gender disparities in patterns of disciplinary action suggest that certain aspects of antisocial behavior may be differentially interpreted by adults based on student characteristics.

3. Self-Efficacy.

Definitions. Self-efficacy (Bandura, 1986) refers to the psycho-social beliefs people have about themselves in their ability to succeed in a task or subject. The CCSR categorizes self-efficacy as an important academic mindset that impacts the degree to which students persevere in academic work and exhibit positive academic behaviors. Individuals tend to engage in activities that they feel confident in their ability to complete and to avoid those in which they lack such confidence (Bandura, 1986). Self-efficacy is distinct from the notion of self-concept, which assesses how people feel about their past performance in relation to others. Self-efficacy measures expectations about whether or not people can successfully perform a specific task at a later point in time. That is, self-efficacy focuses on performance capabilities rather than on personal qualities (Zimmerman, 2000). Self-efficacy is thought to be part of the academic mindsets category of the CCSR framework (Farrington et al., 2012).

Related outcomes. Self-efficacy has been linked to achievement, academic persistence, motivation, and affective factors, among other outcomes.

Achievement. Self-efficacy is central to Schunk's (1987, 2015) theory on how young children acquire cognitive skills. Schunk posited that children's past academic experiences shape their self-efficacy, or how they expect to do on an academic task or in an academic domain. This sense of efficacy is thought to influence students' motivation, which affects their performance. This performance further informs children's sense of self-efficacy, continuing the cycle again. The link between self-efficacy and achievement has been widely explored. A meta-analysis conducted by Multon, Brown, and Lent (1991) summarized the relationship between self-efficacy and academic performance and persistence. Overall, correlational studies suggested that there was a correlation with self-efficacy equal to an effect size of +0.38 for academic performance, and +0.34 for persistence (defined as time spent on task, number of items completed, or number of academic terms completed).

Schunk (1981) tested the effects of a cognitive modeling approach in improving students' math achievement compared to a traditional didactic approach. He found that the cognitive modeling approach increased students' math achievement only through improved self efficacy.

In fact, controlling for previous math achievement, self-efficacy explained almost a quarter of the score on a later mathematics test. Children with higher reported self-efficacy persisted longer on and performed better on arithmetic tasks than students with low self-efficacy.

Similarly, Di Giunta et al. (2013) used structural equation modeling to map relationships among conscientiousness, openness, self-esteem, academic self-efficacy, and achievement among high school students. Controlling for gender, parents' education, and previous achievement, analyses revealed that academic self-efficacy mediated personality traits' and self-esteem's relationship with achievement. Drawing on social cognitive theory, the authors asserted that academic self-efficacy positively predicted achievement during the senior year after accounting for student background characteristics and prior achievement.

Using a sample of Australian students from the 2003 Programme of International Student Assessment (PISA) cohort, Parker, Marsh, Ciarrochi, Marshall, and Abduljabbar (2014) measured the relationship between self-efficacy and achievement, as measured by a tertiary entrance rank (TER), a figure that combines school-based achievement and state-wide standardized testing scores. Latent path modeling revealed that self-efficacy was a positive predictor of TER after controlling for earlier achievement and socioeconomic status. That is, self-efficacy measured at age 15 explained a significant amount of the variance in TER achievement four years later. Such studies provide some evidence that self-efficacy may predict greater academic persistence and higher achievement.

Motivation and affect. Motivation and affect are key to Bandura's (1993) conception of self-efficacy. Bandura (1993) posited that self-efficacy affects cognitive development through four major processes: cognitive, motivational, affective, and selection processes. Put another way, self-efficacy can influence cognitive processes such as decision making and skill utilization, it can affect students' motivation, it can help with affective aspects such as by ameliorating achievement anxiety or depression, and it can influence students' choice of activities and environments. Thus, one of Bandura's (1993) key assertions regarding self-efficacy beliefs was that "people's level of motivation, affective states, and actions are based more on what they believe than on what is objectively true" (p. 2). These processes further shape students' cognitive development and functioning. Bandura (1993) theorized that students with higher self-efficacy participate in classroom activities more readily, work harder, persist longer, and have fewer negative emotional reactions when encountering challenging situations. In particular, students' sense of efficacy can have emotional consequences by decreasing their stress, anxiety, and depression (Bandura, 1993; Schwarzer & Fuchs, 1996).

In sum, correlational studies have shown that self-efficacy is associated with positive outcomes including psychosocial functioning in children and adolescents (Holden, Moncher, Schinke, & Barker, 1990), and higher academic achievement and greater persistence (Multon, Brown, & Lent; 1991; Richardson, Abraham, & Bond, 2012). Gutman and Schoon (2013) suggest that self-efficacy beliefs are essential precursors to both positive cognitive and non-cognitive growth. This is because when children believe they are capable of succeeding in a task, they may be more motivated and emotionally willing to persist, which is likely to further influence their performance on the task. On the other hand, students may perform poorly despite their skills because of the lack of perceived personal efficacy to make optimal use of such skills (Bandura, 1993).

It is important to note that scholars have observed that self-efficacy is often subject-specific (Lennon, 2010). While it can be predictive of more global academic achievement, some have argued that self-efficacy beliefs about specific academic problems best predict improvements in these specified areas or domains (Bong, 1998; Pajares & Miller, 1995; Parker et al., 2014). Accordingly, in the study reported in this manuscript, I focus on reading self-efficacy as the intervention under study has a predominant reading focus.

Part 2. Approaches to Improving Non-Cognitive Factors

Both research and common sense underscore the importance of non-cognitive factors as a desired outcome for students. From a practitioner's standpoint, perhaps more important than the proposed relationships between certain variables is information on what can be done to improve these outcomes for students. As a result, many have proposed ways in which schools can enhance non-cognitive outcomes.

There are a number of ways in which researchers have approached improvements in non-cognitive factors. Dusenbury, Calin, Domitrovich, and Weissberg (2015) summarize four main approaches. One is to incorporate explicit skills instruction in non-cognitive factors, such as teaching students about different kinds of emotions or how to cope with stress. Another approach is to integrate non-cognitive factors with academic curriculum. These programs incorporate non-cognitive factors such as learning empathy within the context of the academic subject. Other programs aim to improve teacher instructional practice to improve teacher-student relationships and a positive classroom climate as a means to improve student outcomes. Finally, whole school approaches involve systematic changes that affect multiple stakeholders in the school such as teachers, parents, and school leaders. This study focuses on this last approach.

While whole school reforms, or comprehensive school reform models, saw great popularity in the 1990s (Borman, Hewes, Overman, & Brown, 2003; Slavin, 2008), renewed interest in such models has emerged in recent times as a result of their potential role in improving non-cognitive as well as academic factors. Research suggests that instead of an isolated intervention or stand-alone curriculum on non-cognitive factors, a whole school approach which involves coordinated changes in instructional management, practices, and relationships among stakeholders may be effective in improving non-cognitive factors (Castrechini & London 2012; Cook, Murphy, & Hunt, 2000; Garcia, 2016).

Why might whole school approaches be more effective than targeted classroom interventions? Ecological systems theory (Bronfenbrenner & Morris, 1998) offers a framework for conceptualizing how an intervention involving multiple levels of a child's environment can affect children's academic and non-academic outcomes. It posits that there are multiple environmental influences that impact students' learning, including the home, classroom, school, and community. The theory suggests children are at the center of a layered system of environments, which have direct, indirect, and interactive influences on their development. As a result, education reforms that involve not only the classroom environment but also the broader school, family, and community contexts are likely to impact students' development more holistically and effectively.

Empirical research on evidence-based programs designed to improve non-cognitive outcomes has also affirmed the effectiveness of whole school approaches. The Collaborative for Social and Emotional Learning released a guide to effective social and emotional learning programs for elementary school students (Weissberg, Goren, Domitrovich, & Dusenbury, 2013), which met stringent methodological criteria. Among the 19 SEL programs that qualified as showing evidence of effectiveness on students' non-cognitive outcomes, 16 included school-wide components, and 17 included both school-wide and parent components. That is, programs that have been found effective have adopted an approach that reaches farther than simply the classroom. One of the most widely disseminated of these whole school approaches is the Success For All (SFA) program.

Part 3. Success For All.

Development of the program.

Success For All began in 1986 as an approach to ensure the success of every child, especially in schools composed predominantly of disadvantaged students. A pilot program in a Baltimore city elementary school proved successful, and since then the program has been adopted by approximately 1000 schools nationwide.

Components of Success For All

Quint et al. (2013) summarize the major components of Success For All in the following way:

1) Reading instruction that is characterized by an emphasis on phonics for beginning readers and comprehension for students at all levels, a highly structured curriculum, an emphasis on cooperative learning, across-grade ability grouping and periodic regrouping, frequent assessments, and tutoring for students who need extra help
2) Whole school improvement components that address non-instructional issues
3) Strategies to secure teacher buy-in, provide school personnel with initial and ongoing training, and foster shared school leadership (p. iii)

The following examines each component in more detail.

Reading instruction. Reading Roots is a reading, writing, and language arts program taught to beginning readers in SFA (Slavin, Madden & Chambers, 2009). The Reading Roots curriculum is based on a cooperative learning program that incorporates reading, writing, and language arts. Reading Roots focuses on building students' comprehension and thinking skills, fluency, and liking of reading through activities such as independent reading, direct instruction, and story-related activities. It offers a fast-paced, structured curriculum that teachers are encouraged to follow consistently. An important component of the Reading Roots curriculum is that students are regrouped across grade lines, according to students' reading levels. Within classes students are assigned to 4-5 member heterogeneous teams that differ by performance, sex, and age. The teams sit together and earn points based on the teams' average quiz scores and completion of a variety of assignments. Teams have a responsibility to make sure every team member is learning and achieving.

Students are given reading comprehension assessments every eight weeks to check for understanding. Those who have difficulty with reading are assigned tutors. SFA tutors are usually certified teachers or highly trained para-professionals who work one on one with children who are experiencing the most difficulty in learning for twenty minutes each day.

Whole school improvement components. In addition to the structured reading program, a significant component of the SFA model is student support through family and community resources. Slavin et al. (2009) explain that the Success For All program places a strong emphasis on improving the school's ability to relate to parents and involve them as well as health and social service agencies in addressing students' nonacademic problems. Family Support Teams aim to support student success through four main activities. First, they promote parent involvement through participation in school governance, in-school volunteer support, frequent teacher and school communication to parents, and parent involvement in the curriculum at home. Second, Family Support Teams strive to improve school attendance through a monitoring system and interventions aimed at setting a school-wide norm for regular attendance. Third, a school-based intervention identifies students who are struggling with serious family, behavioral, or attendance problems and refers them to the Family Support Team. The Team reviews the case carefully and creates an action plan accordingly. Finally, the Family Support Team integrates the community with the school. Local community services or social agencies are called upon to assist with specific student needs. For instance, if a family is without heat or shelter, the Family Support Team can connect them with housing assistance programs or other relevant services. Thus, the family-school-community partnerships that are fostered by the Family Support Teams are aimed at improving circumstances that may help or hinder student learning outside the classroom.

Another specific component of the whole school approach is the Getting Along Together (GAT) curriculum. GAT is a school-wide social problem solving curriculum. It teaches students to solve interpersonal problems in a peaceful and productive way. GAT is further composed of three components. First, in a set of classroom lessons called Learn About It, teachers key social problem solving skills in the first two weeks of the school year. Second, Think It Through sheets guide students to reflect and use self-talk as a way to develop prosocial decision making skills. Third, a step-by-step problem solving model called Peace Path helps students resolve conflicts.

The Peace Path discussions, which are held at a roundtable or at a weekly class council meeting, guide students to express their feelings, listen to others, and suggest and agree on a solution.

Strategies to facilitate change process. Because SFA is a comprehensive program with many parts, facilitating the change process within newly adopting schools is important. There are a number of supports to help with this transition. The program requires 80% of teachers' votes at the schools as a part of the schools' application process. In addition to teacher buy-in, extensive professional development and coaching is provided. Initial training typically takes 3 days in the summer before implementation followed by monthly coaching, but training is continued intermittently throughout the first year. A school-based facilitator, who usually is a very experienced teacher, helps to maintain quality implementation of the program. The facilitator offers coaching to teachers through classroom visits and follow-up discussions, manages coordination so that teachers, tutors, family support staff, and other personnel are coordinating their efforts in effective ways, holds component team meetings twice a month to address specific issues that teachers and tutors encounter, and uses data from assessments to individualize support for struggling students.

The program recognizes that in order for effective change to occur, all school staff must work together towards a common, well-defined goal. As a result, school leadership teams are supported with a detailed action plan, professional training, set targets and progress reports, and a culture of achievement. The leadership teams are supported so that they can take responsibility in reaching their school-wide goals.

Program effects of SFA on non-cognitive factors

As the above components illustrate, SFA is not simply a stand-alone curriculum or a limited change in instructional practice; it instead entails the coordination of multiple stakeholders inside and outside of the school to ensure that students are successful academically as well as socially. It is neither practical nor the aim of this study to pinpoint which components of the program relate to specific outcomes. Rather, it views Success For All as a holistic school improvement process that works as a system and culture to improve student outcomes. While the primary purpose of SFA is to improve reading instruction and outcomes, the program includes various non-instructional components such as student social support, a Getting Along Together program, a solutions team focused on behavior, attendance, and other non-academic outcomes, strong professional development, and support in school leadership. Given these holistic supports, SFA may also have an impact on students' non-cognitive factors in addition to academic outcomes.

Only one study to date has explored the non-cognitive impacts of Success For All (Muñoz & Dosset, 2004). Other evaluations have collected data on student problem behavior (Jones, Gottfredson, & Gottfredson, 1997) or retention and attention (Madden et al. 1993) but the measures were not designed to enable comparison with control groups or differences over time. Muñoz and Dosset (2004) evaluated the program among three treatment and three matched control groups in Kentucky. In addition to reading outcomes, they also investigated the program's impact on the number of disciplinary actions (out-of-school suspensions), absences, and perceptions of school climate, educational quality, and teacher satisfaction as supplementary analyses. Although the authors did not mention the term "non-cognitive factors" specifically in the study, out-of-school suspensions and absences are often viewed as proxies for behavior issues. The results of the study revealed that SFA students scored significantly higher on reading measures with an effect size of +0.11. Moreover, SFA schools saw nearly double the growth in attendance compared to control schools. Reductions in out-of-school suspensions varied year by year, but overall, treatment schools showed a notable decrease in school suspensions after the three-year period compared to control schools. However, the authors note only simple differences in gains and do not adjust for covariates or establish comparability of treatment and control groups in their analyses of these non-cognitive outcomes. Despite these methodological limitations, the results of this study suggest the promise that SFA has in enhancing not only academic outcomes but also non-cognitive outcomes such as fewer suspensions and greater attendance.

Review of literature on similar whole school reform programs

While the literature on SFA and non-cognitive outcomes is currently limited to the one study mentioned above, evaluations of other similar whole school reform programs have evidenced positive impacts on comparable outcomes. The following explores three related models as illustrative examples.

Comer's School Development Program. Several experimental studies have suggested the effectiveness of Comer's School Development Program for social and emotional outcomes. This program aims to improve interpersonal relationships and the school's social climate in a school as a way to enhance academic outcomes. Three structures help accomplish the goal: 1) a School Planning and Management Team that consists of teachers, administrators, parents, and sometimes students, 2) a Social Support Team, consisting of welfare-counselors, nurses, social workers, and special education teachers, and 3) a Parent Team consisting of parents who support the school with governance, fundraising, volunteering and so forth. These teams work in concert to develop a school improvement plan together that focuses not only on improving academics but also ensuring students' psychological, physical, and social well-being.

A quasi-experimental study of Comer's School Development Program (Cook, Murphy, & Hunt, 2000) involved 19 inner city Chicago schools. Results of multilevel model analyses revealed that students in treatment schools had better non-cognitive outcomes. The program had small positive effects on self-efficacy in school, positive expectations for life, and a decrease in negative social behaviors such as acting out. Students' academic outcomes also improved as a result of the School Development Program.

Positive Action. Another program that has been found effective is Positive Action (e.g. Flay, 2001, Flay, 2003). The goal of the program is to promote positive actions in the physical, intellectual, emotional and social areas, to prevent negative behaviors, and to improve school performance. The program seeks to accomplish this goal through a 6-unit content taught in K-12 curriculum on social and character development, a school-wide climate development program, a counseling program, family program, and community program. Bavarian et al. (2013) carried out a randomized control trial with 14 schools in low-income Chicago schools and reported that Positive Action schools had higher teacher ratings of student academic motivation, lower absenteeism, and better reading scores than those in control schools. Similarly, Flay, Acock, Vuchinich, and Beets (2006) conducted a randomized trial evaluation of the program in Hawaii. It revealed that Positive Action schools had better performance, reduced negative behaviors such as absenteeism and suspensions, and better student-reported attitudes toward positive behaviors.

Finally, Snyder et al. (2010) also conducted a randomized control trial of the program with a different group of schools in Hawaii and found that Positive Action school reported lower absenteeism, fewer suspensions, and fewer grade retentions than control schools.

SWPBS. The School-wide Positive Behavior Support (SWPBS) program is a universal prevention strategy that has no set curriculum but instead aims to alter the school's organizational context to reduce student behavior problems and improve academics. It implements a 3- tiered prevention framework to target students with differing intervention needs and clearly articulates positive behavioral expectations throughout the school. In a randomized, wait-list control effectiveness trial with schools from Illinois and Hawaii, Horner et al. (2009) evaluated differences in perceived school safety, level of problem behavior, and academic achievement. The results suggested that schools implementing SWPBS were perceived as safer environments by students and provided preliminary evidence that SWPBS schools had lower office discipline referrals. Office discipline referrals represents a measure of problem behavior, a construct related to antisocial behavior in this study.

Collectively, the select programs and evaluations reviewed above illustrate that whole school reforms, those that do not involve isolated interventions but multiple components to improve the context of the school, can be effective for students' non-cognitive outcomes as well as achievement outcomes. These whole school programs share similarities with SFA in their comprehensive and inclusive approach to reorganization of the school. Thus, there is good reason to believe that SFA may support students' non-cognitive outcomes in addition to their academic performance. Unfortunately, with the exception of one study (Muñoz & Dosset, 2004), there has been no research that explores such a hypothesis with a broader range of non-cognitive factors.

This is particularly problematic because the current policy context underscores the need for rigorous evaluations of programs. ESSA requires states to have at least one measure of non-academic factors such as student engagement or school climate, but it also urges states to implement programs that are evidence based. According to ESSA, an evidence based intervention is described as one that:

> (i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on –
>
> > (I) strong evidence from at least one well-designed and well-implemented experimental study;
> >
> > (II) moderate evidence from at least one well-designed and well-implemented quasi-experimental study; or
> >
> > (III) promising evidence from at least one well-designed and well-implemented correlational study with statistical controls for selection bias

Instead of implementing programs or interventions that appear to make sense on face value or based on personal experience, states are urged to ensure that the programs are supported by well-designed and implemented studies that have strong theoretical bases. Thus, there is a more urgent need for high-quality research that supports the effectiveness of promising programs, especially in the area of improving non-cognitive factors.

Part 4. Relationship Between Achievement and Non-Cognitive Factors

Another issue in the non-cognitive literature is that there is a lack of nuance in the literature linking non-cognitive outcomes and achievement. While many assume from a theoretical standpoint that the two are reciprocally related (Farrington et al., 2012), empirical studies linking the two concepts have most often assumed that non-cognitive factors cause achievement or treated them as separate outcomes (Durlak et al., 2011; Gutman & Schoon, 2012). For instance, in a review of SEL studies, Corcoran, Cheung, Kim, & Xie (2017) reported on the effects of SEL programs on academic achievement and found an effect size of +0.25 in reading, +0.27 in math, and +0.19 in science. Studies such as these operate under the premise that SEL or non-cognitive programming will improve student achievement. Other assessments of SEL programs treat SEL outcomes and academic outcomes as separate domains (e.g., Bavarian et al. 2013; Bradshaw, Mitchell, & Leaf, 2010). Fewer studies have examined the alternative that achievement affects non-cognitive outcomes, yet there is some evidence that this directional relationship exists as well.

For instance, in a longitudinal study of low-income children, Miles and Stipek (2006) found that low academic achievement predicted later aggression among students. More specifically, path analyses indicated that poor literacy achievement in the first and third grades predicted greater aggressive behavior in the third and fifth grades, respectively. They explained that "[c]hildren who have difficulty learning to read . . . may become frustrated or unhappy in school and express their frustration and unhappiness by acting aggressively toward the teacher or classmates" (p. 104). Similarly, Jimerson and Ferguson's (2007) longitudinal study indicated that grade retention in early grades due to poor academic achievement was linked to greater aggressive behaviors by grade 8 compared to those who were not retained. Morgan, Farkas, Tufis, and Sperling (2008) explored whether early achievement in first grade would predict behavior problems in the third grade. Using multilevel logistic modeling to analyze the ECLS-K data, the authors found that struggling readers in the first grade were more likely to be rated as having poor task engagement, poor self-control, externalizing behavior problems, and internalizing behavior problems by the third grade. This was true even when controlling for earlier ratings of each of the behavior outcomes, socioeconomic background, and other demographic variables.

This relationship with achievement has been found in other non-cognitive domains as well. Finn and Cox (1992) found that in a longitudinal sample of elementary students whether a student was highly engaged or not was linked to academic achievement, even as early as in the first grade. They further found that academic achievement measured at grades 1 and 3 positively related to student engagement at grade 4, suggesting that early academic achievement may predict later engagement.

Similarly, Shouse, Schneider, and Plank (1992) used the NELS: 88 data predicted that students' academic ability (as measured by standardized test score) and performance (as measured by teacher afforded grades) would determine teacher rated academic engagement. They predicted that students in the lowest test quartile would be rated lower on academic engagement since they may be frustrated and disenchanted with school because of their poor performance or come from families that cannot support their academic pursuits as well. Indeed, Shouse et al. (1992) found that students' academic achievement was a predictor of engagement and that significant differences by race and school type were also found.

It has been long established that students who learn reading skills and improve their strategy use are likely to perform better in reading (Bandura, 1997). When students perceive this increase in performance, they are likely to develop higher self-efficacy for reading. In a similar vein, students who are academically struggling are more likely to be disengaged in school (Willms, 2003). Bloom (1976) summarized this notion well:

> At the other extreme are the bottom third of students who have been given consistent evidence of their inadequacy…over a period of five to ten years. Such students rarely secure any positive reinforcement in the classroom… from teachers or parents. We would expect such students to be infected with emotional difficulties [and to] exhibit symptoms of acute distress and alienation from the world of school and adults. (Bloom, 1976, p. 158)

There is a body of literature that suggests that achievement affects non-cognitive outcomes and another body of literature that suggests non-cognitive outcomes predict achievement. However, studies linking these two relationships simultaneously are sparse.

The Possible Mediating Relationship Between Achievement and Non-Cognitive Factors

There is a related gap in the literature around *how* whole school reform programs affect non-cognitive outcomes. That is, many programs have a theoretical basis for how program components change students' non-cognitive skills, behaviors, and attitudes (e.g. Battistich, Schaps, & Wilson, 2004; Conduct Problems Prevention Research Group, 1999; Flay & Allred, 2003), but there is little empirical investigation of these pathways. An exploration of the mechanisms through which programs affect change in individual outcomes is warranted to better understand the processes through which programs impact student outcomes.

In the context of this study, how might SFA influence non-cognitive outcomes? Based on the literature reviewed previously, SFA may accomplish this through academic achievement.

Since SFA is designed with a very strong academic program, it is possible that students' increased academic performance may lead to better non-cognitive outcomes. The figure below graphically depicts this possibility. The figure illustrates that Success For All's whole school components may lead to improved non-cognitive outcomes (direct effect), but that it may also achieve this through improved student achievement (mediated effect).

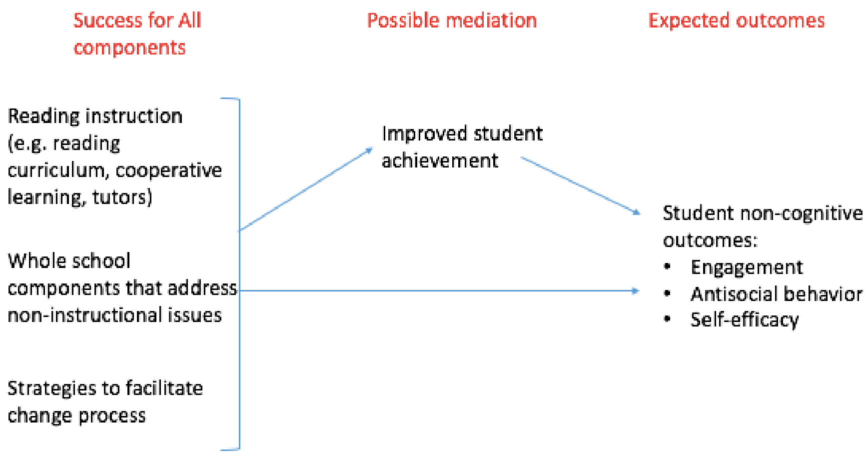

Figure 2. Possible mediation between Success For All and non-cognitive outcomes through achievement.

A relatively long line of research has indicated that SFA meets rigorous standards for effectiveness. Borman et al. (2007) evaluated thirty-five urban Midwestern and rural Southern elementary in a randomized control trial. After three years of implementation, multilevel analyses of approximately 2,100 students found significant impacts on three subtests of the Woodcock Reading Mastery Tests – Revised (WMTR), with effect sizes ranging from +0.21 to +0.36 for post-test Word Attack scores.

A report from the Investing in Innovation (i3) evaluation of an SFA scale-up project was conducted by a third-party evaluator (Quint et al., 2015). Thirty-seven evaluation schools in five school districts were randomly assigned to SFA or a control group. After three years of implementation, students in SFA scored significantly higher in phonics skills. Second graders who had started kindergarten in the bottom half on literacy skills scored significantly higher on measures of phonics skills, word recognition, and reading fluency than similar control group students.

In 2016, Success For All was one of thirty programs invited to the What Works Showcase at the White House, a platform for sharing education programs that are supported by rigorous evidence. The What Works Clearinghouse (2017) lists nine studies evaluating SFA that met its standards for evidence of effectiveness. The expected change in percentile rank for an average comparison group student if he or she had received the SFA program was 9 units for alphabetics and 12 units for reading fluency.

While the literature on SFA's effect on achievement is abundant, there are also some studies as aforementioned that give rise to the idea that the program may also affect the non-cognitive domain (Madden et al. 1993; Muñoz & Dosset, 2004; Jones, Gottfredson, & Gottfredson, 1997). Based on the broader literature that suggests that non-cognitive outcomes influence academic achievement and vice versa, it is possible that achievement is the link between programs and non-cognitive factors, acting as a mediator. That is, students in SFA may achieve more academically than students who are not in the program, which then leads to higher non-cognitive factors.

To summarize, the research on SFA confirms the effectiveness of the program in improving student achievement. In addition to academic achievement being an important outcome of interest, it may also function as a mediator to further enhance students' non-cognitive factors. This is because, as reviewed previously, achievement has been found to predict non-cognitive factors. There is increasing interest in promoting non-cognitive factors, and it is important to understand whether improving students' achievement is one way to achieve this goal.

Summary

In summary, a long history of research has underlined the importance of non-cognitive factors in student success in school and beyond. The notion of evaluating students' non-cognitive factors in addition to academic achievement has gained traction in recent years. Currently, some of the most effective approaches to improving students' non-cognitive outcomes have involved whole school reform models. One of the most widely used whole school reform models in the United States is SFA. While SFA has been repeatedly shown to positively impact academic outcomes, the non-academic components of the program and its similarity to other whole school programs that have been found to be effective with non-cognitive factors make it reasonable to hypothesize that there may be a program effect on such outcomes as well. In particular, this study focuses on engagement, antisocial behavior, and reading self-efficacy as non-cognitive outcomes of interest. Indeed, given the scale of SFA that is used across schools in the U.S. and the importance of evidence-based evaluations of programs, there is a need to examine whether SFA is effective in improving these non-cognitive outcomes for students.

In addition to investigating the relationship between SFA and non-cognitive factors, there is also a gap in the literature regarding the relationship between academic achievement and non-cognitive outcomes. The vast majority of evaluations of school-based programs report program effects on achievement and non-cognitive factors as separate outcomes or suggest that non-cognitive factors impact achievement (e.g. Flay, Acock, Vuchinich, & Beets, 2006; Freiberg et al., 2011; Jones, Brown, Hoglund, & Aber, 2010). While this may be true, it is also possible that increased achievement may lead to increased non-cognitive skills. As such, this study seeks to clarify the relationship between non-cognitive factors and achievement over time which will contribute to the broader literature that links the two concepts. Finally, if the relationship between achievement and non-cognitive outcomes is established, there is a possibility that SFA may work through achievement as a mediator to impact non-cognitive outcomes.

By exploring potential program effects, relationships between achievement and non-cognitive factors, and a hypothesized mechanism through which changes in non-cognitive outcomes may be attained, this study seeks to address the current gaps in the literature and extend knowledge on effective programs and how they work to improve student outcomes.

Chapter 3. Methods

Research Questions

The proposed study seeks the answer the following research questions with subsequent hypotheses:

1. Are there differences in non-cognitive factors among students in SFA and non-SFA schools over time?

   H1: It is hypothesized that children in SFA will score higher on non-cognitive factors.

2. To what extent does participation in SFA relate to growth in non-cognitive factors over time?

   H2: It is hypothesized that students participating in SFA will have higher rates of growth in non-cognitive factors compared to non-SFA students.

3. To what extent does achievement predict non-cognitive outcomes over time, regardless of SFA status?

   H3: It is hypothesized that achievement will positively predict non-cognitive outcomes.

4. To what extent do non-cognitive factors predict achievement over time, regardless of SFA status?

   H4: It is hypothesized that non-cognitive factors will positively predict achievement.

5. To what extent is the effect of SFA on students' non-cognitive factors mediated by academic achievement?

   H5: It is hypothesized that SFA schools will have higher ratings of academic achievement, which in turn will positively impact students' non-cognitive factors.

Procedure

Data used in the current study come from the Study of Instructional Improvement (SII), conducted by the University of Michigan and the Consortium for Policy Research in Education, which was a longitudinal, quasi-experimental study evaluating the effectiveness of three different whole school reform models (America's Choice, Accelerated Schools, and Success For All) compared to control schools in improving achievement. The original SII data included 115 schools (roughly 30 schools in each of the three interventions under study, plus 26 matched control schools). Schools were selected for the study in four steps. First, a list was drawn of all U.S. public elementary schools that had participated in one of the three comprehensive school reform models in the 1998-1999, 1999-2000, or 2000-2001 school years. Secondly, a set of 17 geographic regions was selected from which to sample schools. Thirdly, intervention schools from the 17 geographical regions were selected with an attempt to balance the samples of schools from the intervention programs by matching on length of participation in the intervention program and socioeconomic disadvantage. The sampling procedure deliberately called for an oversampling of high-poverty elementary schools in order to understand instructional improvement in high-poverty settings (Correnti, 2007). In the last step, comparison schools were chosen from within the same 17 geographical regions. Additionally, comparison schools were selected so that their distribution on a disadvantage index matched that of intervention program schools. Because America's Choice and Accelerated Schools are no longer in use today, the analytical focus is on comparing Success For All schools with control schools only.

Data were collected during the 2000-2001 through 2003-2004 academic years as two cohorts of students were followed simultaneously – Cohort A following students from kindergarten through second grade and Cohort B following students from third grade through fifth grade. In this study, I focus only on Cohort A students for two reasons. The first is that research has focused attention on the younger grades in particular because findings suggest that "early skills gaps, both cognitive and non-cognitive, translate into differences in students' subsequent learning and development" (Garcia, 2015, p. 9). That is, children with stronger non-cognitive skills at school entry are more likely to succeed academically than those with weaker initial skills. As Heckman observed, "skills beget skills" suggesting that strong learning skills acquired early on facilitate skills in the future (Heckman, 2008). Thus, early investments in education is likely to set a more favorable pathway toward adolescent and adult development (Heckman, 2008; Heckman & Kautz, 2012; Cunha & Heckman, 2007). The second reason is that because the study does not differentiate among schools that have implemented the whole school reform programs for one year or for many years, it is difficult to establish an equivalent baseline among SFA schools since the start year may have been different for each school. Consequently, some third grade students may have been in an SFA school for two years while third graders at a different SFA school may be in their first year of the program. This problem is not present for the K-2 sample since kindergarten is the first year of school across all participating schools and establish an equivalent starting point in terms of years enrolled in an SFA school.

Data were collected in two waves in a staggered design. One cohort of students' data was collected in the 2000-2001 year when students entered as kindergartners through 2002-2003 when they were second graders. The second cohort of students' data was collected in the 2001-2002 school year when students entered as kindergartners and through 2003-2004 when they were second graders. If students left the school or were not available for data collection, they were replaced by students who entered the study at the beginning of the academic year. For this study, I do not distinguish between the data collection start year and consider all students who were part of the 3-year longitudinal sample, not replacement students.

Analytic Sample

There were 977 students (SFA=469, control=508) students within 54 schools (SFA =29, control = 25) who had outcome data at kindergarten in the first year of data collection. By the second year, there were 923 students who had data (SFA = 443, control = 480), and 897 students by the end of the third year (SFA= 425, control = 472). No schools were dropped and there was no differential attrition between treatment and control groups. A comparison of the analytic sample with the nationally representative ECLS data shows that there were differences in characteristics between the analytic sample and the rest of the US. Table 1 compares the characteristics of the analytic sample with ECLS data for illustrative purposes to provide a snapshot of how nationally represent the analytic sample may be. Schools in the current study represented a higher proportion of African-American children (38.2%) compared to ECLS (15.7%). Whites make up the largest ethnic group in the ECLS sample (57.3%), while the SII sample includes less than half that percentage (27.3%). Moreover, the ECLS sample came from families who generally had a higher total income than those in the analytic sample. Students in the analytic sample also came from a higher proportion of families that received food stamps (26.4%) compared to the ECLS sample (19.8%). In all, the comparison suggests that students in the current study come from more disadvantaged backgrounds compared to the national sample. This is expected since the sampling procedure for the original study purposely oversampled high-poverty schools.

Table 2 shows the demographic characteristics of the analytic sample by treatment and control group. A series of chi-square tests revealed that the treatment and control schools had students with similar backgrounds. Only race was significantly different between the two groups (χ^2 = 4.47, p< 0.034) while the groups were similar on all other measures. Thus, there is good reason to believe students were equivalent at baseline on individual and family characteristics.

Despite the efforts to match treatment and control schools, there was not a perfect match in terms of school characteristics. Table 3 shows that SFA schools within the study had more disadvantaged backgrounds based on covariates used in the study. For instance, SFA schools included a higher percentage of schools who had parents on welfare than control schools (17% vs. 14%) and larger proportions of students receiving free lunch (71% vs 64%).

Table 1.

Current study/ECLS sample demographic comparison

	Current study (n=977)	ECLS (weighted n =3,865,797)
Demographics		
Male	50.7%	51.3%
Female	49.3%	48.7%
White	27.3%	57.3%
Black	38.2%	15.7%
Hispanic	24.1%	19.3%
Reported Total Family Income		
UNDER $5,000	3.2%	3.4%
$5,000 - $9,999	7.3%	5.0%
$10,000 - $14,999	8.3%	7.8%
$15,000 - $19,999	8.7%	6.9%
$20,000 - $24,999	7.3%	7.7%
$25,000 - $29,999	9.2%	6.3%
$30,000 - $34,999	6.2%	7.0%
$35,000 - $39,999	5.1%	5.5%
$40,000 - $49,999	8.0%	10.3%
$50,000 - $74,999	16.26%	20.0%
$75,000 - $99,999	4.2%	9.5%
$100,000 - $199,999	3.0%	8.7%
$200,000 or more	0.1%	1.9%
Family Received Public Assistance		

AFDC/TANF received in last 12 months	16.5%	12.0%
Foodstamps received in last 12 months	26.4%	19.8%

Table 2.

Comparison between SFA and control students on general demographics

	SFA (n=469)	Control (n=508)	p
Gender			
Male	53.5%	48.0%	
Female	46.5%	52.0%	
Race			
White	27.3%	27.4%	
Black	42.9%	33.9%	**
Hispanic	19.6%	28.2%	**
Other	10.2%	10.6%	
Reported Total Family Income			
UNDER $5,000	3.1%	3.3%	
$5,000 - $9,999	5.6%	8.8%	
$10,000 - $14,999	8.7%	8.0%	
$15,000 - $19,999	8.5%	8.8%	
$20,000 - $24,999	6.3%	8.2%	
$25,000 - $29,999	7.7%	10.6%	
$30,000 - $34,999	7.0%	5.5%	
$35,000 - $39,999	5.1%	5.1%	
$40,000 - $49,999	8.0%	8.0%	
$50,000 - $74,999	18.4%	14.4%	
$75,000 - $99,999	4.3%	4.0%	
$100,000 - $199,999	2.4%	3.53%	
$200,000 or more	0.0%	0.2%	
Family Received Public Assistance			
AFDC/TANF received in last 12 months	16.6%	16.3%	
Food stamps received in last 12 months	26.1%	26.7%	

Table 3.

Comparison between treatment and control schools on covariates

	Mean in treated (N=29)	Mean in untreated (N=25)	Standardized difference
Proportion on welfare	0.19	0.14	0.38
Census disadvantage index	1.08	0.72	0.26
Proportion individuals who do not have high school diploma	0.35	0.32	0.18
Proportion single parent households	0.16	0.14	0.23
Proportion individuals unemployed	0.12	0.11	0.13
Locale	2.17	1.52	0.81*
Percent free lunch	0.73	0.65	0.44
Total school enrollment	456.76	513.48	-0.33
Percent minority	0.81	0.74	0.25

Propensity score matching. Because there was imbalance of covariates even after initial matching of schools, propensity score methods were used to balance the school data since the treatment was at school level.

There are several benefits of propensity score matching methods compared to traditional regression adjustment methods. First, propensity score matching approximates randomization because it attempts to balance treatment and control groups based on a set of observable covariates (Kee, 2012). It can thus strengthen causal inferences about the impact of the SFA program since it attempts to create a better match between treatment and control groups based on observed covariates and attempts to eliminate selection bias. Selection bias refers to the phenomenon in which those who decide to participate in the study may be systematically different from those who do not. It should be noted, however, that propensity score matching methods cannot control for unobserved differences and thus cannot eliminate issues of selection bias completely. Second, the approach involves a single summary index of a group of covariates simultaneously, so it is more efficient and less computationally demanding for subsequent analyses (Becker & Ichino, 2002). Third, propensity score matching requires no assumption about the functional form of the relationship between outcomes and predictors of outcome unlike parametric techniques that often assume a linear or sometimes nonlinear (e.g. logistic) functional form.

There are a number of different propensity score matching methods, the three most common of which are one-to-one matching, stratification, and weighting (Austin, 2011). One-to-one matching works best when the sample size is large since those units without a good match are not included in the analysis. Given the relatively small sample size of schools in this study, this method was not preferred. Between stratification and weighting methods, stratification was selected for this study because it retained all units and provided the best balance of covariates in preliminary analyses.

In propensity stratification, propensity scores are first estimated from logistic regression predicting assignment to the treatment group based on a number of covariates. For this study, the equation to calculate propensity scores for each school was as follows:

$$\pi_i = \mathrm{P}(\mathrm{T}_i = 1 | \mathrm{X}_\mathrm{i}),$$

where π_i is the propensity score for school i , which is the conditional probability (P) of assigning a school to treatment group (T = 1) give a set of covariates (X) of school i. The covariates were selected based on the recent report by Cheung & Slavin (2016) using the same data and included the ten covariates in Table 3.

In the second step, students are divided into strata based on the propensity score, and treatment effects are observed within strata. Quintiles are often used for adjustment, since they are expected to remove 90% of the confounding (Austin, 2011; Cochran, 1968). For analysis, propensity score stratification quintiles were added to adjust for differences within strata. Table 4 shows that once propensity score stratification was applied, balance of covariates between treatment and control schools was greatly improved. There were no significant differences in the covariates below once propensity score stratification was applied.

Table 4.

Comparison between treatment and control schools with propensity strata

	Mean in treated (N=29)	Mean in untreated (N=25)	Standardized difference
Proportion on welfare	0.19	0.18	0.06
Census disadvantage index	1.08	1.04	0.03
Proportion individuals who do not have high school diploma	0.35	0.35	0.03
Proportion single parent households	0.16	0.16	0.03
Proportion individuals unemployed	0.12	0.12	0.05
Locale	2.17	2.15	0.03
Percent free lunch	0.73	0.71	0.11
Total school enrollment	456.76	465.61	-0.05
Percent minority	0.81	0.78	0.11

Data Collection Instruments.

The following instruments were used to collect the data.

The Student Motivation Form. The Student Motivation Form (SMF) asked students how they perceive themselves in academic interests or skills and was administered each spring. The survey asked children to report on how much they enjoy reading and mathematics, how easy or hard reading and mathematics are for them, and any behaviors with which they might struggle that may also interfere with their learning. For kindergarten through second grade students, the SMF was administered individually in an easel format, and an assessor recorded the students' answers on a single-sided scan answer form. The response rates for this form ranged from 96-97% across the three years for all study participants.

Student Rating Form. Mathematics and language arts teachers were asked to complete a Student Rating Form (SRF) for each student. Most students had only one teacher as the sample were younger students who do not have departmentalized classes, but for those who had more than one form completed, the ratings were averaged. The SRF instrument gathered information on a student's academic engagement, approaches to learning, and problem behaviors (if any). The response rates for this form ranged from 89-92% across the three years for all study participants.

Achievement. The TerraNova, a standardized achievement test, was administered to all students in the fall and spring. Kindergarten students took the Letter/Word identification sections of the Woodcock-Johnson Tests of Achievement – Revised as a pretest in the fall but completed the TerraNova assessment in the fall and through the end of second grade. The TerraNova is a nationally recognized assessment instrument. Table 5 illustrates the data collection schedule for instruments of most relevance to this study.

Table 5.

Data collection schedule for achievement and non-cognitive instruments

Year 1		Year 2		Year 3	
Fall	Spring	Fall	Spring	Fall	Spring
Woodcock Johnson	TerraNova reading	TerraNova reading	TerraNova reading	TerraNova reading	TerraNova reading
	Student Motivation Form, Student Rating Form		Student Motivation Form, Student Rating Form		Student Motivation Form, Student Rating Form

Instructional logs. Data on literacy and mathematics instruction were gathered from separate logs for Language Arts and Mathematics that recorded information about a single day of instruction for a single student. It assessed the amount of emphasis given to important topics in each subject. Because there have been extensive studies examining the use of instructional logs in school improvement using this data (e.g. Hill, Rowan, & Ball, 2005; Hill, Schilling, & Ball, 2004), the instructional logs are not analyzed in the present study except as a proxy to measure fidelity of implementation.

The Parent Survey. The Parent Survey consisted of interviews with parents in the spring whose children were participating in the study and asked questions about basic demographic information and questions about the family's access to basic needs. This provided the basis for the construction of the socioeconomic status measure. The response rates for this form ranged from 62-85% across the three years for all study participants.

School Characteristic Inventory. Multiple sources were combined to generate school-level information. The Quality Education Data (QED) database (a commercially available database) the NCES Common Core Database (CCD), the School Characteristics Inventory, and Parent Survey data responses aggregated to the school level. The school principal mainly completed the School Characteristics Inventory (SCI) questionnaire which was composed primarily of closed-ended questions designed to gather information about the school such as enrollment, funding and programs, and student and staff demographics. Each school was given one SCI to complete. The response rates for this form ranged from 68-100% across the three years for all study participants.

Measures

Each of the outcome measures below were recorded in the spring of each data collection year. The first point of measurement did not occur before the intervention took place (i.e., the fall); because the study attempts to approximate randomization for study schools, the first point of measurement is considered the first post-test.

Engagement. Two measures of student engagement were obtained from teacher and student reports. For the teacher completed measure, eleven items on the Student Rating Form asked about students' engagement behaviors. Sample items include "this student usually pays attention in class" and "this student is eager to learn." The scores ranged from 1 (not at all true) to 4 (very true). The teacher-reported measure of student engagement demonstrated high internal consistency ($\propto$=0.96). A confirmatory factor analysis (CFA) was run to see if the data would load onto one factor as theorized. Fit indices chosen to regard acceptability of the model were: the Comparative Fit Index exceeds .93 (Byrne, 1994), RMSEA less than .08 (Browne & Cudeck,

1993) and a TLI exceeding 0.90 (e.g., Hu & Bentler, 1999). Preliminary analysis revealed that the data fit the one factor structure well ($\chi 2(44, N=948) = 481.83$, $p < 0.01$, CFI=0.95, RMSEA = 0.06, TLI = 0.94), supporting its construct validity.

Seven items on the Student Motivation Form indicate students' engagement difficulties in reading class. Sample items include "It's hard for me to finish my work in reading" and "It's hard for me to pay attention in math class." The scores ranged from 1 (not at all true) to 4 (very true). This measure was reverse coded to attain consistency across measures. That is, students with higher scores would represent those with fewer engagement difficulties. The student-reported measure of engagement demonstrated acceptable internal consistency ($\propto$=0.79). A confirmatory factor analysis (CFA) was run to see if the data would load onto one factor as theorized. Preliminary analysis revealed that the data fit the one factor structure well ($\chi 2(20, N=952) = 133.18$, $p < 0.01$, CFI=0.93, RMSEA = 0.08, TLI = 0.91), supporting its construct validity.

Reading self-efficacy. Efficacy in reading is derived from four items from the Student Motivation Form. Although the original scale consisted of seven items, only four were selected for the reading self-efficacy scale for this study based on face validity of items that reflected the construct of self-efficacy as well as preliminary exploratory factor analysis. Sample items include "I do well in reading" and "I learn things quickly in reading." The scores ranged from 1 (not at all true) to 4 (very true). The scale demonstrated a Cronbach's alpha of 0.60, which was lower than the reliabilities of the other outcome measures. Preliminary analysis revealed that the data fit the one factor structure well ($\chi 2(2, N=972) = 13.98$, $p < 0.01$, CFI=0.97, RMSEA = 0.08, TLI = 0.90), supporting its construct validity.

Antisocial behavior. A measure of disruptive behaviors was derived from eight items from the Student Rating Form. Sample items include "this student often acts impulsively" and "this student disrupts the work of others." The scores ranged from 1 (not at all true) to 4 (very true). This measure was reverse coded to attain consistency across measures. That is, higher scores represent students with less antisocial behavior. The teacher-reported measure of antisocial behavior demonstrated high internal consistency ($\propto= 0.93$) Preliminary analysis revealed that the data fit the one factor structure well ($\chi2(14, N=940) = 157.36$, $p < 0.01$, CFI=0.95, RMSEA = 0.06, TLI = 0.94), supporting its construct validity.

Achievement. The TerraNova scale scores for reading, which can range in value from 100 to 900, were used in subsequent analyses. Achievement was aggregated at the school level for use as a covariate in multilevel analyses and as a mediator in the mediation analysis. Kindergarten students took the Letter/Word identification sections of the Woodcock-Johnson Tests of Achievement – Revised as a pretest in the fall; this measure was standardized in the original SII study.

School-level information. SII used the Quality Education Data (QED) database (a commercially available database), the NCES Common Core Database (CCD), a School Characteristics Inventory, and Parent Survey data responses aggregated to the school level. Multiple measures of school context were included for the construction of propensity scores: Proportion of families on welfare, census disadvantage index, proportion individuals who do not have high school diploma, proportion single parent households, proportion of individuals unemployed, locale, percent free lunch students, total school enrollment, and percent minority students.

Student-level information. Student race, gender, and grade information gathered from school records were used as covariates. Race was coded into four categories: White (reference category), Black, Hispanic, and other. Student SES was derived from parent surveys. The original study used the mean average of five-item measures: the highest education levels reported for the (1) mother and (2) father, (3) reported total family income level, and the occupational prestige scores of the (4) mother and (5) father to produce a standardized coefficient of SES representing a student's status compared to other students in the study population.

Normality of variables. A preliminary step in the analysis explored whether the data met the normality assumption. If variables violate the assumption of normality, parameter estimates may be biased (Hong, Yoo, You, & Wu, 2010). Using West, Finch, and Curran's (1995) widely used guidelines of normality (skewness <2, kurtosis <7), all the continuous study variables met the assumption of normality.

Implementation Fidelity

The original study unfortunately did not systematically record implementation fidelity of SFA or other whole school reform programs. However, based on instructional logs that SFA and control teachers submitted, it was possible to determine whether teachers were following the design for reading that is outlined by the SFA program. Using data from the original study, Rowan and Miller (2007) found that Success For All schools, which were designed to have more programmed approaches to instructional change, could be effectively distinguished from the conventional reading instructional practices of control schools. This suggests that SFA schools were implementing the whole school model as planned, at least in terms of the reading instruction component.

Missing Data

Missing data in independent variables and covariates ranged from 0-15% missing across years. Missingness was substantial also for the outcome variables at subsequent time points. As mentioned previously, there were 977 students (SFA=469, control=508) students within 54 schools (SFA =29, control = 25) who had outcome data at kindergarten in the first year of data collection. In the second year, 552 students (SFA=239, control = 313) of the same group of students had outcome data in the spring. In the third year, 369 students (SFA=172, control = 197) of the same group had outcome data in the spring.

Missingness that differs by student characteristics can severely bias estimates (Hausman & Wise, 1979). Accordingly, two missing data analyses were conducted. First, differences in baseline student characteristics and outcomes data between those treatment and control students who had the full data and those who had at least one point of data missing were compared. A series of t-tests and chi-square analyses revealed that there were no significant differences between those with complete data and those with missing data in terms of SES, race, sex, or any of the first year non-cognitive outcome variables. However, students with complete outcome data were more likely to have higher baseline Woodcock-Johnson reading scores ($t(763)=-2.04$), $p<.05$) than those who did not have complete data. The second analysis was to assess whether there were differential rates of missing data between those in the control and treatment groups. A chi-square test suggested that differential missingness was non-significant. That is, the proportion of students who had missing data and those who did not was similar across treatment and control groups.

Power analysis was not done on the final sample because the practice is not recommended to be done post-hoc (Levine & Ensom, 2001; O'Keefe, 2007). Instead, confidence intervals are supplied in analyses when appropriate.

Missing data is handled using Full Information Likelihood (FIML) estimation in Mplus. Unlike multiple imputation, FIML does not impute any missing data but rather estimates parameters using all the information that is available from the data set, making it a more efficient estimator (Enders, 2001). Dong and Peng (2013) found that using principled missing data methods such as multiple imputation or full information maximum likelihood produced less biased estimates compared to list-wise deletion even with 60% missing data.

Data analytic strategy

Research question 1. To answer the first research question, two-level multilevel models were conducted using Mplus. This methodological approach allows for the comparison of non-cognitive outcomes between control and treatment students over a three-year time period. Because the data is nested, multilevel models are more appropriate for analysis than regular OLS regression, which does not take into account the clustered nature and can lead to biased error terms and inflated Type I error rates (Raudenbush & Bryk, 2002). Additionally, multilevel models allow for partitioning of variance in student non-cognitive outcomes into both within and between levels (in this case, schools).

Unconditional models were run first as a preliminary step to determine the amount of variance between and within schools.

Level 1:

$Y_{ij} = \beta_{0j} + r_{ij}$

Level 2:

$\beta_{0j} = u_{0j}$

The intraclass correlation (ICC) measures the variation in outcomes between schools and were calculated by the following equation:

$$\text{ICC} = \tau_{00}/(\tau_{00} + \sigma^2)$$

with τ_{00} representing level 2 variance and σ^2 representing level 1 variance. The unconditional models revealed that ICCs were relatively low, ranging from 0.02-0.09 across non-cognitive outcomes. This suggests that the majority of variance lies within schools and not between.

Researchers often use the ICCs along with the average cluster size to compute a design effect to inform whether analyses that accounts for clustering should be used. The design effect represents the degree to which standard errors are underestimated in a complex sample (such as one that involves clustering) compared to a simple random sample (Maas & Hox, 2005). It is calculated as:

$$\text{deff} = 1 + (c - 1) \times \text{ICC}.$$

where c represents the average cluster size (Muthén, & Satorra, 1995). According to Lai and Kwok (2015), a design effect of 1.1 or higher warrants multilevel modeling when researchers are interested in the effects of higher level predictors, as substantially biased standard errors will be estimated when analyzed at the single level. The design effects in this study ranged from 1.3 to 2.4, suggesting that multilevel modeling is appropriate.

Next, the following two-level models were run to address research question 1:

Level 1:

$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Gender}) + \beta_{2j}(\text{Race}) + \beta_{3j}(SES) + \beta_{4j}(Noncognitive\ outcome_{t-1}) + r_{ij}$

Level 2:

$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Mean achievement pretest}) + \gamma_{02}(\text{treatment}) + \gamma_{03}(\text{propensity strata}) + u_{0j}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

$\beta_{3j} = \gamma_{30}$

$\beta_{4j} = \gamma_{40}$

This represents the spring non-cognitive outcome for student *i* in school *j* regressed on the Level 1 residual variance, r_{ij}. This analysis was repeated for each of the four outcome variables. Students' gender, race, and SES, and previous year non-cognitive outcome were also added as predictors in the Level 1 model. Gender, race, SES, and previous year non-cognitive outcomes were considered fixed slopes because preliminary analyses suggested that variances of random slopes were not significant. At Level 2 of the model, SFA treatment effects were estimated on each mean spring non-cognitive outcome in school *j*. School-level covariates included the school mean pretest achievement score that year to help reduce the unexplained variance in the outcome and propensity strata dummy coded into quintiles to account for differences between treatment and control schools. This analysis was run for each year of the three-year longitudinal sample, similar to the cross-sectional analysis done for the longitudinal sample in Cheung & Slavin's (2016) article. Three separate analyses were done for each of the outcomes (self-reported engagement, teacher-reported engagement, anti-social behavior, and self-efficacy). One model was run with all students who had outcome data in kindergarten (Year 1), another model with students who remained the following year and had outcome data through the first grade (Year 2), and one model with students who remained all three years and had outcome data through the second grade (Year 3). Effect sizes were calculated by dividing the coefficient with the raw standard deviation of the outcome variable (Becker, 1998; Feingold, 2009). For the longitudinal multilevel models, the raw standard deviation of the outcome variable at baseline was used, following recommendations by Feingold (2009). This represents the mean difference in relation to the standard deviation of the outcome variable.

Research question 2. To answer research question 2, the following three-level unconditional model was run first to determine ICCs in a growth curve model:

Level 1:

$Y_{tij} = \pi_{0ij} + e_{tij}$

Level 2:

$\pi_{0ij} = \beta_{00j} + r_{0i}$

Level 3:

$\beta_{00j} = \gamma_{000} + u_{00j}$

where π_{0ij} is the mean outcome between time points within student, β_{00j} is the mean outcome between students, and γ_{000} is the mean outcome between schools. ICCs ranged from 0.06-0.08 between schools and from 0.87-0.89 between time points within individuals.

Next, predictors were added to each level of the model for each of the non-cognitive outcomes:

Level 1:

$Y_{tij} = \pi_{0ij} + \pi_{1ij}$(time) + π_{2ij}(Achievement pretest) + e_{tij}

Level 2:

$\pi_{0ij} = \beta_{00j} + \beta_{1q}$(Gender)+ β_{2q}(Race) +β_{3q}(SES) + r_{0i}

$\pi_{1ij} = \beta_{10j}$

$\pi_{2ij} = \beta_{20j}$

Level 3:

$\beta_{00j} = \gamma_{000} + \gamma_{001}$(treatment)+ γ_{002}(propensity strata) + u_{00j}

$\beta_{10j} = \gamma_{100} + \gamma_{110}$(treatment)

$\beta_{20j} = \gamma_{200}$

where π_{0ij} represents the average initial outcome for student *i* at time *t*= 0, which in this case is the first year of data collection, controlling for reading achievement pretest. The covariates included gender, race, and SES, which were fixed variables at level 2. Treatment status and propensity strata were fixed covariates at level 3. The key difference in the equations is that treatment was included as a predictor of the time slope at level 3, creating a cross-level interaction. This interaction term would indicate whether the treatment had a relationship with the rate of growth or decline in non-cognitive outcomes. Again, this analysis was run separately for each of the four outcome variables (self-reported engagement, teacher-reported engagement, anti-social behavior, and self-efficacy). Because time is specified within the first level, this model includes all students regardless of whether they had outcome data at all time points, using FIML again to account for missing data. This is necessary to model trends over time, regardless of whether all students had data for the three time points or not. The key difference between the previous model and the model under discussion was that the longitudinal model takes into account the rate of change over time whereas the previous model looks cross-sectionally at whether there are differences in treatment and control students' non-cognitive outcomes for each additional year they were part of the study.

Research question 3-5. To answer research questions 3 through 5, auto-regressive cross-lagged (ARCL) panel models were used. ARCL models are based on structural equation modeling and used to examine the structural relations of repeatedly measured constructs (Selig & Little, 2012). The "auto-regressive" part of the term describes the stability of individual differences from one measurement occasion of the variable to the next. The "cross-lagged" part of the term describes the effect of a variable on a different variable at a later occasion. Such an approach is advantageous in observing relationships over time and to simultaneously address reciprocal influences (Cacioppo, Hawkley, & Thisted, 2005).

Moreover, although both traditional multilevel models and SEM models of the sort described above can address direct and indirect effects in nested data (Curran, 2003), the SEM approach was favored for two main reasons in examining mediation in particular. First, traditional multilevel models can produce confounding and erroneous conclusions when assessing mediation because such mediation conflates between-group and within-group effects (Zhang, Zyphur, & Preacher, 2009). Second, the SEM approach described above emphasizes the longitudinal, change over time process and is thus more intuitive when considering the development of mediation effects over time.

The ARCL method allows for stronger inference about the direction of paths compared to cross-sectional analyses because it implies that each construct is a function of the same construct at a previous time in addition to some random disturbance component (Selig & Preacher, 2009). In other words, the temporal precedence of one variable before another can be suggestive of a causal relationship (Selig & Little, 2012).

Panel models are useful for mediation purposes in particular because they allow estimation of the direct and indirect effects while taking into account different measurement errors of variables over time. They present an advantage over cross-sectional mediation models, as Maxwell, Cole, and Mitchell (2011) argue that substantial bias can occur in estimating both total and partial mediation with cross-sectional analyses in longitudinal data. The authors explain that the cross-sectional models are generally misspecified when used to estimate longitudinal processes because it does not take into account the effects of the independent variable (i.e., X) on the mediator (i.e., M) and the outcome (i.e., Y) and of mediator on the outcome over time. It further assumes that X at a certain time causes M at the same time point, precluding the possibility that causation can happen at different time points (Maxwell, Cole, & Mitchell, 2011).

Baseline ARCL Model. In the context of this study, the ARCL model to test mediation paths between treatment status and non-cognitive outcomes through achievement was as follows, where M represents the mediating variable (student achievement) and Y represents the non-cognitive outcome of interest (i.e., teacher-reported antisocial behavior, teacher-reported engagement, student-reported engagement, and student-reported reading self-efficacy). The general equation is as follows:

$$M_t = \beta_{M,[t-1]} M_{[t-1]} + \beta_X X_t + \beta_{Y,[t-1]} Y_{[t-1]} + \zeta_{M,[t]}$$

$$Y_t = \beta_{y,[t-1]} Y_{[t-1]} + \beta_{M,[t-1]} M_{[t-1]} + \beta_X X_t + \zeta_{Y,[t]}$$

where Y_t represents the measure of the outcome variable Y at time t (ranging from 1 to 3), and M_t represents the measure of the mediating variable M for individuals at time t (ranging from 1 to 6). $\zeta_{M,[t]}$ and $\zeta_{Y,[t]}$ are the residuals for individuals for the mediating and outcomes variables, respectively. $\beta_{M,[t-1]} M_{[t-1]}$ and $\beta_{y,[t-1]} Y_{[t-1]}$ represent the autoregressive parameters for the mediating and outcome variables, respectively. $\beta_{Y,[t-1]} Y_{[t-1]}$ and $\beta_{M,[t-1]} M_{[t-1,]}$ are the cross-lagged coefficients for the mediating and outcome variables, respectively. These parameters represent the prediction of one construct at time t from the other construct at previous time points, controlling for autoregressive predictions of each construct (Hong et al., 2010). Appropriate control variables were added to each pathway. Figure 3 below illustrates the proposed model.

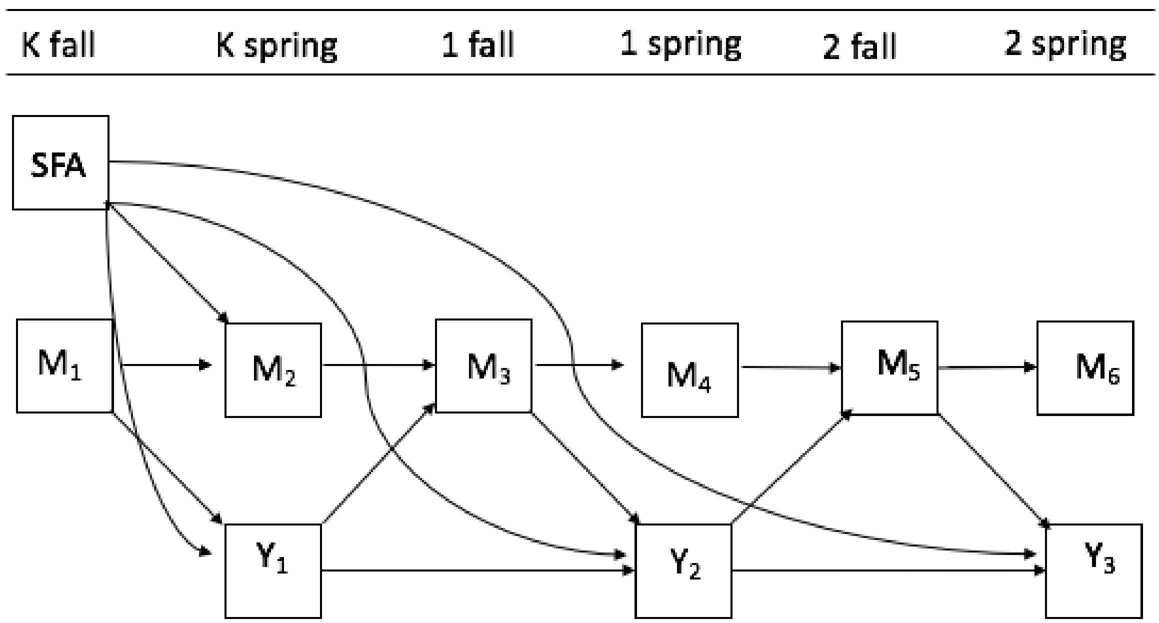

Figure 3. Model 1: Proposed ARCL model with M representing achievement and Y representing non-cognitive outcomes.
Note. School level and student level control variables are accounted for in this model.

A number of observations should be noted with this model. First, relations only one lag apart were specified. Second, the temporal precedence of the mediator in relation to the outcome variable was maintained. For instance, it was not hypothesized that Y_2 would impact M_3 in the reverse direction, to keep in line with the longitudinal ordering of variables. Third, measurement errors (not pictured in the diagram for simplicity) were considered correlated across the time points for each indicator, in line with the literature that suggests that measurement errors of a repeated measure may covary (Ma & Xu, 2003; Pitts, West, & Tein, 1996). Fourth, SFA was hypothesized to have direct effects on the outcome variables as well as the mediating variable at the first time point with the assumption that SFA's effects on achievement in the first year would impact achievement in subsequent years.

The direct effect of SFA on Y_1 , Y_2, and Y_3 are parameters of interest as well as the mediated effect of SFA on these outcomes through the mediated pathways. The two mediated pathways in this model are $SFA \rightarrow M_2 \rightarrow M_3 \rightarrow Y_2$ and SFA $\rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \rightarrow M_5 \rightarrow Y_3$.

ARCL Models Exploring Additional Specifications. However, the mediation paths described above may not be the only mediated pathways possible. To explore possible nested models, two more specifications were explored. First, additional models added in direct effects from SFA to the mediating outcomes at each time point, accounting for the possibility that SFA had a different effect on achievement at different time points (Cheung & Slavin, 2016). Second, additional models included the effect of the mediator two lags or higher on the outcome variable such that M_2 could also predict Y_2 and M_3 and M_4 could also predict Y_3, for instance.

This approach of starting with a simpler model and exploring additional specifications was chosen because MacKinnon (2012) suggests that a series of nested models be used to test hypotheses in model building. He suggests that "a simple model could be used with comparison of nested models used to decide the parameters to include in the longitudinal mediation model" (p. 208). While it is also possible to start with more specifications and compare with simpler models, other researchers have used the build-up approach espoused in this study to determine the best fitting model (e.g. Bentley, 2011; Pardini, Loeber, & Stouthamer-Loeber, 2005). Thus, although both model building approaches have been used in the literature, the approach to add specifications to a simpler model was chosen to more clearly distinguish whether these additional direct effects or lagged effects improved the model.

In order to determine which model is favorable, assessment of model-data-fit was accomplished through examination of standard model fit indices (e.g. RMSEA, CFI, NNFI) to achieve the best fitting model. Standardized parameters were calculated to represent effect sizes.

In order to adjust for the nested nature of the data, robust standard errors were calculated. Clustered errors, which result from non-independence of observations, cause the standard estimator for the variance to be biased downward (Cameron & Miller, 2015). One of the ways to correct for these clustered errors is to calculate robust standard errors. Robust standard errors can be calculated using Huber/Pseudo ML/sandwich corrections that are robust to non-normality and non-independence of observations, allowing for correlation among observations. Thus, using such estimators accounted for the nested nature of data in the ARCL models while also taking advantage of the models' ability to observe direct and mediated effects over time.

Chapter 4. Results

Baseline statistics

Table 6 provides descriptive statistics of variables at baseline for treatment and control groups as well as means for first year outcome variables. As noted in Chapter 3, a series of chi-square tests revealed that the treatment and control schools had students with similar demographic backgrounds. Only race was significantly different between the two groups (χ^2 = 4.47, p< 0.034), while the groups were similar on all other measures. Treatment students were more likely to be Black than control students, while there was a larger proportion of Hispanic students in the control group.

Table 6.

Descriptive statistics

	Control			Treatment		
	Mean (SD) / %	Range	n	Mean (SD)	Range	n
Student level						
Male	48.03		244	53.52		251
Female	51.97		264	46.48		218
White	27.36		139	27.29		128
Black	33.86		172	42.86		201
Hispanic	28.15		143	19.62		92
Other	10.63		54	10.23		48
SES	-0.02 (0.87)	-1.52-3.37	453	0.07 (0.80)	-1.52-2.49	414
Teacher-reported engagement (first year)	2.96(0.68)	1-4	508	3.06(0.64)	1-4	469
Student-reported engagement difficulties (first year)	2.51 (0.84)	1-4	508	2.47 (0.76)	1-4	469

Student-reported self-efficacy (first year)	3.30 (0.70)	1-4	508	3.26 (0.68)	1-4	469
Teacher-reported antisocial behavior (first year)	3.19 (0.70)	1-4	508	3.20 (0.70)	1-4	469
Woodcock-Johnson reading comprehension scale (pretest)	0.09 (1.03)	-2.49-2.99	414	-0.11 (1.03)	-2.35-2.99	394
School level						
Community disadvantage index	0.72 (1.32)	-1.02-4.07	25	1.08 (1.50)	-0.62-4.30	29
Proportion households with assistance income	0.14 (0.11)	0.-0.43	25	0.19 (0.14)	0.03-0.49	29
Proportion individuals without high school diploma	0.32 (0.14)	0.07-0.64	25	0.35 (0.18)	0.09-0.71	29
Proportion single-parent households	0.14 (0.11)	0.03-0.56	25	0.16 (0.08)	0.04-0.32	29
Proportion individuals unemployed	0.11 (0.07)	0.01-0.33	25	0.12 (0.08)	0.02-0.31	29
Locale						
Large City	56		14	27.59		8
Midsize city	9		36	34.48		10
Urban Fringe of Large City	8		2	31.03		9
Urban Fringe of Mid-size City	0		0	6.90		2
Percent free lunch	0.65 (0.21)	0.22-0.96	25	0.73 (0.19)	0.28-0.95	29

Total school enrollment	513.48 (199.7)	247-1052	25	456.76 (135.52)	291-892	29
Percent minority students	0.75 (0.28)	0.15-1	25	0.81 (0.28)	0.16-1	29
Woodcock-Johnson reading comprehension scale	0.13 (0.39)	-0.39-1.03	25	-0.14 (0.36)	-0.90-0.55	29

Bivariate correlations between each of the key student-level baseline variables (binary or continuous) and first year outcome variables are presented in Table 7. Bivariate correlations between the outcome variables over time are presented in Appendix B.

First year teacher-reported engagement engagement was positively correlated with sex and SES. Student-reported engagement difficulties (reverse coded) was positively correlated with SES and teacher-reported engagement. Reading self-efficacy was positively correlated with teacher-reported engagement; antisocial behavior (reverse coded) was positively correlated with sex and teacher-reported engagement. Finally, Woodock-Johnson scores were positively correlated with SES, teacher-reported engagement, student-reported engagement difficulties (reverse coded), reading self-efficacy, and reading self-efficacy. Multicollinearity was not detected as variation inflation factors (VIFs) values were low.

The student- and school-level variables (binary or continuous) were separated into two correlation matrices for ease of interpretation. Among the school level covariates, the proportion of welfare status was positively correlated with the census disadvantage index; proportion of parents with no high school diploma was positively correlated with census disadvantage index and welfare status. The proportion of single parent households was positively correlated with census disadvantage index, welfare status, and proportion of parents with no high school diploma. Percent of households unemployed was also positively correlated with all the above variables along with proportion of single parent households. Percent free and reduced price lunch was positively correlated with census disadvantage index, proportion of families on welfare, those without high school diplomas, single parents, and those who are unemployed. Finally, the proportion of minority students in the school was positively correlated with census disadvantage tract, proportion of parents without high school diplomas, proportion of families unemployed, and total school enrollment. Multicollinearity was detected only for the census disadvantage index, but was not deemed problematic as variables with high VIFs are acceptable if used as control variables (Allison, 2012).

Table 7.

Correlational analysis among study variables

Student level variables

	1.	2.	3.	4.	5.	6.	7.
1. Sex (1=male, 0=female)	1.00						
2. SES	-0.07	1.00					
3. Teacher-reported engagement	0.19**	0.16**	1.00				
4. Student-reported engagement difficulties	0.05	0.13**	0.11*	1.00			
5. Reading self-efficacy	0.01	0.01	0.13**	-0.05	1.00		
6. Antisocial behavior	0.13**	0.07	0.56**	0.09	0.03	1.00	
7. Woodcock-Johnson	0.02	0.28**	0.35**	0.13**	0.13**	0.05	1.00

*$p<.05$ **$p<.01$

School level variables

	1.	2.	3.	4.	5.	6.	7.	8.
1. Community Disadvantage Index	1.00							
2. Proportion on welfare	0.94**	1.00						
3. Proportion with no high school diploma	0.83**	0.73**	1.00					
4. Proportion single parent houe hold	0.78**	0.77**	0.57**	1.00				
5. Proportion unemployed	0.91**	0.80**	0.74**	0.71**	1.00			
6. Percent free and reduced lunch	0.54**	0.48*	0.45*	0.50**	0.58**	-0.26	1.00	
7. Total school enrollment	0.05	-0.06	0.18	0.09	0.06	0.04	-0.15	1.00
8. Percent minority	0.45*	0.35	0.54**	0.34	0.46*	-0.20	0.80**	0.01

$^*p<.05$ $^{**}p<.01$

Multilevel models.

Unconditional model. Unconditional models run for each of the four outcome variables revealed the proportion of variance that lies within the student and school levels (Table 8). Overall, student level variance was large compared to school level variance. Specifically, 2% of variance in reading self-efficacy was between schools, while approximately 9% of variance in teacher- and student-reported engagement and antisocial behavior was between schools.

Table 8.

Variances and intraclass correlations from unconditional models

	Teacher-reported engagement	Student-reported engagement	Reading self-efficacy	Antisocial behavior
Student level variance	0.40 (0.02)	0.59 (0.03)	0.47 (0.02)	0.45 (0.02)
School level variance	0.04 (0.01)	0.06 (0.02)	0.01 (0.01)	0.04 (0.01)
Intraclass correlation (ICC)	0.09	0.09	0.02	0.09

Note. All variances were significant at the p<.05 level.

Cross-sectional multilevel models. To address research question 1, a series of cross-sectional multilevel models were run for each additional year of the study students participated in. Tables 9 through 12 display the regression coefficients for each outcome variable for each year.

The results indicate that only teacher-reported engagement was significantly different between treatment and control groups, with kindergarten students in SFA scoring higher (b=0.18, ES=+0.27) than those in comparison schools.

While this was the only significant treatment effect that was detected, there were other notable relationships among the outcome variables and student-level covariates. For instance, male (b=-0.27, ES=-0.41 in first year), Black (b=-0.29, ES=-0.45 in the third year), and lower SES students (b=0.14, ES=+0.21 in the first year) were likely to have lower teacher-reported engagement. Previous average reading scores were positively associated (b=0.004, ES=+0.01 in the second year) with higher teacher-reported engagement but to a small degree. Similarly, male (b=-0.10, ES=-0.13 in the first year), Black (b=-0.26, ES=-0.24 in the third year), and lower SES (b=0.10, ES=+0.13 in the first year) students were likely to have lower self-reported engagement. For reading self-efficacy, Hispanic (b=-0.23, ES=-0.35 in the third year) and lower

SES students (b=0.06, ES=0.10 in the third year) were likely to report lower self-efficacy. Finally, male (b=-0.19, ES=-0.27), lower SES (b=0.10, ES=0.24 in the second year), and all races other than White students were likely to have lower ratings of antisocial behavior.

Table 9.

Cross-sectional multilevel model predicting teacher-reported engagement

	Year 1 (kindergarten) (n=977)	Year 2 (first grade) (n=923)	Year 3 (second grade) (n=897)
Intercept	3.09 (0.07)**	-0.85 (0.94)	2.11(0.86)*
Student level fixed effects			
Male	-0.27(0.04)**	-0.18(0.04)**	-0.21(0.07)**
SES	0.14(0.02)**	0.05(0.03)✝	0.07(0.03)*
Black	-0.09(0.06)	0.00(0.07)	-0.29(0.12)*
Hispanic	0.08(0.07)	0.03(0.08)	-0.22(0.13)✝
Other	0.08(0.08)	0.15(0.08)✝	0.05(0.15)
Previous measure of teacher-reported engagement		0.57(0.04)**	0.30(0.07)**
School level fixed effects			
Treatment	0.18(0.08)*	-0.12(0.07)	0.04(0.06)
Previous average score on language test	0.02(0.12)	0.00(0.00)*	0.00(0.00)
Propensity strata 2	0.11(0.1)	-0.01(0.09)	0.03(0.07)
Propensity strata 3	-0.06(0.11)	0.22(0.07)**	0.01(0.08)
Propensity strata 4	-0.08(0.09)	0.1(0.09)	0.11(0.09)
Propensity strata 5	-0.13(0.08)	0.23(0.09)**	0.06(0.09)
Random components			
School level variance	0.03 (0.01)**	0.01 (0.01)	0.00(0.03)
Student level variance	0.37 (0.02)**	0.25 (0.02)	0.29(0.04)**

Note. ✝ p<.10, *p<.05 **p<.0. Parameter estimate standard errors listed in parentheses.

Table 10.

Cross-sectional multilevel model predicting student-reported engagement (reverse coded)

	Year 1 (kindergarten) (n=977)	Year 2 (first grade) (n=923)	Year 3 (second grade) (n=897)
Intercept	2.70 (0.11)**	2.28(1.72)	0.87(1.69)
Student level fixed effects			
Male	-0.1(0.05)*	-0.12(0.07)	0.07(0.07)
SES	0.1(0.03)**	0.06(0.05)	0.07(0.06)
Black	-0.18(0.1)	-0.1(0.11)	-0.26(0.11)*
Hispanic	-0.2(0.12)	-0.13(0.13)	-0.2(0.12)
Other	-0.15(0.1)	0.05(0.12)	0.09(0.12)
Previous measure of student-reported engagement		0.19(0.05)**	0.21(0.04)**
School level fixed effects			
Treatment	-0.01(0.09)	0.04(0.12)	-0.04(0.12)
Previous average score on language test	-0.09(0.12)	0(0)	0(0)
Propensity strata 2	-0.03(0.16)	0.07(0.1)	0.17(0.09)
Propensity strata 3	0.1(0.14)	0.15(0.14)	0.2(0.19)
Propensity strata 4	-0.09(0.13)	0.05(0.12)	-0.12(0.14)
Propensity strata 5	-0.13(0.16)	0.05(0.14)	0.09(0.15)
Random components			
School level variance	0.04(0.01)**	0.03 (0.02)	0.00(0.02)
Student level variance	0.58 (0.03)**	0.64(0.04)**	0.54(0.03)**

Note. **p*<.05 ***p*<.0. Parameter estimate standard errors listed in parentheses

Table 11.

Cross-sectional multilevel model predicting student-reported reading self-efficacy

	Year 1 (kindergarten) (n=977)	Year 2 (first grade) (n=923)	Year 3 (second grade) (n=897)
Intercept	3.22(0.11)**	1.97(1.22)	3.73(1.14)**

Student level fixed effects			
Male	-0.01(0.05)	-0.08(0.06)	-0.09(0.06)
SES	0.01(0.03)	-0.04(0.03)	0.06(0.03)*
Black	0.09(0.07)	-0.10(0.07)	-0.11(0.07)
Hispanic	-0.07(0.07)	-0.07(0.09)	-0.23(0.11)*
Other	-0.03(0.1)	-0.14(0.07)*	-0.09(0.11)
Previous measure of reading self-efficacy		0.18(0.04)**	0.11(0.06)*
School level fixed effects			
Treatment	-0.04(0.06)	0.10(0.08)	0.01(0.08)
Previous average score on language test	0.15(0.08)	0.00(0.00)	0.00(0.00)
Propensity strata 2	0.13(0.09)	0.00(0.08)	0.02(0.08)
Propensity strata 3	0.11(0.12)	0.07(0.09)	0.11(0.06)
Propensity strata 4	0.04(0.12)	-0.10(0.10)	0.00(0.12)
Propensity strata 5	0.08(0.12)	-0.06(0.09)	0.00(0.08)
Random components			
School level variance	0.01(0.01)	0.01(0.01)	0.00(0.04)
Student level variance	0.47(0.02)**	0.37(0.04)**	0.35(0.04)

Note. *p<.05 **p<.0. Parameter estimate standard errors listed in parentheses

Table 12.

Cross-sectional multilevel model predicting teacher-reported anti-social behavior (reverse-coded)

	Year 1 (kindergarten) (n=977)	Year 2 (first grade) (n=923)	Year 3 (second grade) (n=897)
Intercept	3.31(0.08)**	-0.23 (1.11)	0.75(0.85)
Student level fixed effects			
Male	-0.19(0.04)**	-0.14(0.04)**	-0.13(0.05)**
SES	0.06(0.03)*	0.10(0.03)**	0(0.04)
Black	-0.25(0.06)**	0.08(0.07)	-0.22(0.08)**
Hispanic	0.03(0.08)	0.19(0.08)*	-0.10(0.09)
Other	0.07(0.07)	0.25(0.08)**	0.03(0.13)

Previous measure of teacher-reported antisocial behavior		0.6(0.04)**	0.57(0.05)**
School level fixed effects			
Treatment	0.08(0.07)	-0.04(0.07)	-0.16(0.09)
Previous average score on language test	0.04(0.13)	0.00(0.00)	0.00(0.00)
Propensity strata 2	0.05(0.1)	0.07(0.10)	0.08(0.08)
Propensity strata 3	0.14(0.11)	0.09(0.10)	0.06(0.08)
Propensity strata 4	-0.1(0.1)	0.11(0.11)	0.33(0.11)**
Propensity strata 5	0(0.13)	0.07(0.11)	0.2(0.12)
Random components			
School level variance	0.43(0.03)**	0.02(0.01)**	0.00(0.01)
Student level variance	0.03(0.01)**	0.25(0.02)**	0.26(0.03)**

Note. *p<.05 **p<.0. Parameter estimate standard errors listed in parentheses

Longitudinal multilevel model. Next, to address research question 2, 3-level growth curve models were run. Results (Table 13) indicated that participation in SFA reached practical significance (b=0.14, p=0.08) in predicting teacher reported engagement, suggesting that SFA students had marginally higher engagement than control students in the spring of year 1. Interaction effects were run to address whether SFA could affect the rate of growth in non-cognitive outcomes (research question 2), but the interaction terms were not significant across all outcomes. Notably, student achievement, a time-varying variable, significantly and positively predicted three out of the four outcomes (teacher-reported engagement, student-reported engagement and reading self-efficacy), while being male predicted lower teacher-reported engagement (b=-0.28, ES=-0.42) and greater antisocial behavior (b=-0.24, ES=0.34, reverse coded). SES positively predicted teacher-reported engagement (b=0.11, ES=+0.17), student-reported engagement (b=0.06, ES=+0.08), and lower antisocial behavior (b=0.08, ES=+0.11, reverse coded). Being Black was linked to lower student-reported engagement (b=-0.16, ES=-0.20) and greater antisocial behavior (b=-0.21, ES=-0.30).

Table 13.

Longitudinal multilevel models

	Teacher-reported engagement	Student-reported engagement	Reading self-efficacy	Antisocial behavior
Intercept	3.05(0.07)**	2.60(0.09)**	3.32 (0.07)**	3.31 (0.08)**
Level 1: Within student				
Student achievement	0.15(0.02)**	0.12(0.02)**	0.10(0.01)**	0.03(0.02)
Level 2: Between student				
Male	-0.28(0.04)**	-0.06(0.04)	-0.04(0.03)	-0.24(0.04)**
SES	0.11(0.02)**	0.06(0.03)*	-0.01(0.02)	0.08(0.03)**
Black	-0.04(0.06)	-0.16(0.07)*	0.04(0.06)	-0.21(0.06)**
Hispanic	0.12(0.07)	-0.17(0.09)	-0.07(0.06)	0.08(0.08)
Other	0.12(0.07)†	-0.02(0.08)		0.1(0.08)
Level 3: School				
Treatment	0.14(0.08)†	-0.02(0.09)	-0.04(0.05)	0.04(0.09)
Propensity strata 2	0.11(0.10)	0.05(0.10)	0.06(0.04)	0.06(0.10)
Propensity strata 3	0.01(0.09)	0.20(0.09)*	0.08(0.05)	0.15(0.11)
Propensity strata 4	-0.03(0.10)	0.02(0.09)	-0.02(0.07)	-0.06(0.12)
Propensity strata 5	-0.03(0.09)	0.02(0.09)	0.03 (0.06)	0.01(0.14)
Treatment x time	-0.03(0.04)	-0.02(0.05)	0.04(0.04)	-0.03(0.04)
Random components				
School level variance	0.02(0.01)**	0.01(0.01)**	0.00(0.00)	0.03(0.01)**
Student level variance	0.15(0.02)**	0.09(0.02)**	0.06(0.01)**	0.24 (0.02)**
Within student variance	0.17(0.01)**	0.5(0.03)**	0.36(0.02)**	0.17(0.01)**

Note. † p<.10, *p<.05 **p<.0. Parameter estimate standard errors listed in parentheses

Autoregressive Cross-Lagged Panel Models

Finally, to address research questions 3 through 5, a series of ARCL panel models were run. SFA represented the treatment variable, achievement measured at various time points represented the mediator, and the outcome variables were non-cognitive factors. Because the multilevel models indicated only teacher-reported engagement as having a significant relationship with SFA, the following ARCL models focused on this outcome only. However, the auto-regressive cross-lagged panel models without the direct effect of SFA for the other non-cognitive outcomes were run for completeness in addressing research questions 3 and 4. The results of these models are included in Appendix C.

The results of the current models on teacher-reported engagement can be found in Tables 15 and 16. Model 1 is based on Figure 4, and represents the most simplistic model with one lag effects and a direct effect from treatment to the spring of kindergarten reading achievement. Model 2 added the possibility of more than one semester lagged effects from reading achievement to teacher-reported engagement. Additional direct effects were also investigated. Model 3 explored a direct effect from treatment to the fall of first grade, while Model 4 added more than one semester lagged effects from mediator to outcome. Similarly, Model 5 explored a direct effect from treatment to the spring of first grade, and Model 6 added the extended lagged effects for this model. Finally, Model 7 explored the possibility of a direct effect from treatment to the fall of second grade. No other lagged effects from reading achievement to teacher-reported engagement were remaining. The models were run as seven separate models instead of one full model with all possible pathways run simultaneously because the number of paths estimated would exceed the number of observations (i.e., clusters) available, causing the model to be under identified and consequently producing unreliable standard errors.

Tables 15 and 16 do not display the control variables or covariances for simplicity, but each model included significant 1) control variables at each direct effect, 2) autocorrelation residuals among the mediating and outcome variables, and 3) covariances between residual terms of downstream variables within the same wave. Non-significant main research question paths were still included; only significant control variables and covariances were included to maintain the most parsimonious model. In these models, only sex, SES, Black, Hispanic, and two of the propensity strata were significant control variables.

Model fit comparison. For the model fit comparisons here, more stringent criteria were used to better distinguish among models. Hu and Bentler (1999) suggest that a cutoff value close to .95 or higher for TLI and CFI, .08 or lower for SRMR, and .06 or lower for RMSEA indicates good fit between the hypothesized model and observed data. Model fit indices (Table 14) indicated that while all models fit the data relatively well, models 5, 6, and 7 showed best fit according to these metrics. The AIC and BIC are also included here as absolute fit indices that do not require nested models. Comparatively, the AIC and BIC favored Models 1 and 2 since they had lower values compared to Models 5, 6, and 7. This contrasts with the other fit indices mentioned previously. Furthermore, chi-square difference tests using the MLR estimator in Mplus and Satorra-Bentler scaling correction did not find significant differences between models with the addition of more than one time-lag relationship from the mediator to outcome and those without (e.g. Model 2 versus Model 1, respectively). This suggests that models with additional time-lagged relationships did not significantly fit the data better than those without. Although collectively there were some discrepancies in the fit indices, the indices more commonly suggested for reporting in structural equation modeling (i.e., RMSEA, CFI, TLI, and SRMR; Hu & Bentler, 1999; Kline, 2005) were used to determine best fit among the models.

Table 14.

Model fit comparisons

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
$\chi 2$/df	120.01/48	119.57/46	158.70/60	156.85/84	58.22/68	58.16/68	71.04/53
RMSEA [CI]	0.04 [0.03, 0.05]	0.04 [0.03, 0.05]	0.04 [0.04, 0.05]	0.04 [0.04, 0.05]	0.02 [0.00, 0.03]	0.02 [0.00, 0.03]	0.02 [0.00, 0.03]

CFI	0.95	0.95	0.93	0.93	0.99	0.99	0.99
TLI	0.92	0.92	0.91	0.91	0.99	0.98	0.99
SRMR	0.06	0.07	0.07	0.07	0.04	0.04	0.04
AIC	11232.83	11233.58	11247.70	11248.79	11729.29	11731.28	11712.40
BIC	11464.83	11316.20	11474.96	11480.78	11943.82	11950.58	11931.70

Diagrams of Model 1 (initial model; Figure 6) and Model 7 (best fitting model; Figure 5) are included. Across all models, there were some common patterns. First, the autoregressive components were significant for both mediator and outcome variables. That is, previous measures of achievement significantly and positively predicted successive measures of achievement. For instance, in Model 7, achievement measured in the spring of first grade had a positive relationship with achievement at the beginning of second grade (β=0.93). Similarly, previous measures of teacher-reported engagement positively predicted later measures of engagement. For instance, in Model 7, engagement measured in first grade had a significant positive relationship with engagement at second grade (β=0.81, 95% CI [0.65,0.97]). Cross-lagged relations were also significant. Previous measures of reading achievement positively predicted later measures of teacher-reported engagement at all years. For instance, in Model 7 achievement at the fall of kindergarten positively predicted teacher-reported engagement at the end of the spring (β=0.40, 95% CI [0.30,0.49]). Previous measures of teacher-reported engagement predicted some later measures of achievement. Specifically, teacher-reported engagement measured at kindergarten positively predicted achievement in first grade in Model 7 (β=0.16, 95% CI [0.05,0.26]), but engagement at first grade did not predict achievement at second grade (β=0.03, 95% CI [-0.06,0.12]).

Moreover, there were no relations between reading achievement and teacher-reported engagement that were significant beyond one lag apart. For instance, achievement in the spring of kindergarten did not predict engagement at second grade in Model 2 (β=0.02, 95% CI [-

0.06,0.10]). Importantly, across all models, no significant mediation effects were detected. Even when considering the combination of different direct effects at varying time points and relationships between teacher-reported engagement and achievement that are more than one lag apart, there was no significant mediation.

As supplementary analysis, reverse mediation was examined. That is, exploratory analysis was done using the same ARCL models, in which engagement mediated the relationship between SFA and later achievement. The results indicated that the pathway from treatment affecting first grade fall achievement through teacher-reported engagement at kindergarten reached marginal significance (β= 0.02, p<0.10). Though this outcome and mediation were not the focus outcomes of this study, this analysis was done for completeness of considering the relationship between engagement and achievement over time. Overall, the results suggest that there was no significant mediation either through engagement or achievement.

Table 15.
ARCL Models 1-4

	Model 1		Model 2		Model 3		Model 4	
	β(SE)	95% CI	β(SE)	95% CI	β(SE)	95% CI	β(SE)	95%CI
Achieve1→Achieve2	0.35(0.04)**	[0.24,0.45]	0.35(0.04)**	[0.25,0.46]	0.39(0.04)**	[0.29,0.50]	0.39(0.04)**	[0.29,0.05]
Achieve2→Achieve3	0.41(0.04)**	[0.30,0.52]	0.41(0.04)**	[0.30,0.52]	0.40(0.04)**	[0.30,0.50]	0.40(0.04)**	[0.30,0.50]
Achieve3→Achieve4	0.92(0.05)**	[0.79,1.05]	0.92(0.05)**	[0.79,1.05]	0.91(0.05)**	[0.78,1.04]	0.91(0.05)**	[0.78,1.04]
Achieve4→Achieve5	0.99(0.05)**	[0.86,1.14]	0.99(0.05)**	[0.86,1.14]	1.02(0.05)**	[0.88,1.15]	1.01(0.05)**	[0.88,1.15]
Achieve5→Achieve6	0.93(0.06)**	[0.78,1.08]	0.93(0.06)**	[0.78,1.08]	0.93(0.06)**	[0.77,1.08]	0.93(0.06)**	[0.77,1.08]
Engage1→ Engage2	0.49(0.03)**	[0.41,0.57]	0.50(0.03)**	[0.41,0.58]	0.49(0.03)**	[0.41,0.58]	0.50(0.03)**	[0.42,0.59]
Engage2 → Engage3	0.85(0.07)**	[0.68,1.02]	0.82(0.07)**	[0.64,1.00]	0.86(0.06)**	[0.71,1.01]	0.81(0.07)**	[0.64,0.99]
Achieve1→Engage1	0.37(0.04)**	[0.27,0.48]	0.39(0.04)**	[0.30,0.49]	0.40(0.04)**	[0.30,0.50]	0.40(0.04)**	[0.30,0.50]
Achieve3→Engage2	0.22(0.03)**	[0.14,0.31]	0.21(0.03)**	[0.13,0.30]	0.23(0.03)**	[0.14,0.31]	0.21(0.04)**	[0.12,0.30]
Achieve5→Engage3	0.08(0.04) †	[-0.03,0.2]	0.21(0.03)**	[-0.04,0.19]	0.09(0.04)*	[-0.02,0.20]	0.07(0.04)	[-.04,0.19]
SFA→Engage1	0.11(0.05)*	[-0.02,0.23]	0.11(0.05)*	[-0.02,0.23]	0.11(0.05)*	[-0.02,0.25]	0.11(0.05)*	[-.03,0.25]
SFA→Engage2	-0.01(0.04)	[-0.11,0.09]	-0.01(0.04)	[-0.11, 0.09]	-0.01(0.04)	[-0.11,0.09]	-0.01(0.04)	[-.11,0.09]
SFA→Engage3	0.06(0.04)	[-0.05,0.17]	0.06(0.04)	[-0.04,0.16]	0.07(0.04)	[-0.05,0.18]	0.07(0.04)	[-.04,0.18]
Engage1→Achieve3	0.15(0.04)**	[0.05,0.26]	0.16(0.04)**	[0.05,0.26]	0.16(0.04)**	[0.07,0.26]	0.16(0.04)**	[0.07,0.26]
Engage2→Achieve5	0.03(0.03)	[-0.06,0.12]	0.03(0.03)	[-0.06,0.12]	0.03(0.03)	[-0.06,0.12]	0.03(0.03)	[-.06,0.12]
SFA→Achieve2	0.04(0.06)	[-0.11,0.18]	0.04(0.06)	[-0.11,0.18]				
Model fit modifications								
SFA→Achieve3					-0.01(0.04)	[-0.12,0.11]	-0.01(0.04)	[-.12,0.11]
Achieve2→Engage2			0.02(0.03)	[-0.06,0.10]				
Achieve2→Engage3			0.07(0.05)	[-0.06,0.21]				
Achieve3→Engage3							0.06(0.06)	[-.02,0.17]
Indirect effects								
SFA→Achieve2→Achieve3 →Engage2	0.00(0.00)	[-0.01,0.02]	0.00(0.01)	[-0.01,0.02]				
SFA→Achieve2→Achieve3→ Achieve4→Achieve5→ Engage3	0.00(0.00)	[0.00,0.00]	0.00(0.01)	[-0.00,0.01]				
SFA→Achieve2→Engage2			0.00(0.01)	[-0.01,0.01]				
SFA→Achieve2→Engage3			0.00(0.00)	[-0.01,0.02]				
SFA→Achieve3→Engage2					0.00(0.01)	[0.03,0.02]	0.00(0.01)	[0.03,0.02]
SFA→Achieve3→Achieve4→ Achieve5→Engage3					-0.00(0.00)	[-0.01,0.01]	0.00(0.00)	[-.01,0.01]
SFA→Achieve3→Engage3							0.00(0.00)	[-.01,0.01]

Table 16.

ARCL Models 5-7

	Model 5		Model 6		Model 7	
	β(SE)	95% CI	β(SE)	95% CI	β(SE)	95% CI
Achieve1→Achieve2	0.39(0.04)**	[0.29,0.49]	0.39(0.04)**	[0.29,0.49]	0.39(0.04)**	[0.29,0.49]
Achieve2→Achieve3	0.41(0.04)**	[0.31,0.52]	0.41(0.04)**	[0.31,0.52]	0.41(0.04)**	[0.30,0.52]
Achieve3→Achieve4	0.91(0.05)**	[0.78,1.04]	0.91(0.05)**	[0.78,1.04]	0.91(0.05)**	[0.78,1.04]
Achieve4→Achieve5	1.01(0.05)**	[0.87,1.15]	1.01(0.05)**	[0.87,1.15]	1.01(0..06)**	[0.87,1.15]
Achieve5→Achieve6	0.93(0.06)**	[0.79,1.08]	0.93(0.06)**	[0.79,1.08]	0.93(0.06)**	[0.78,1.07]
Engage1→ Engage2	0.50(0.03)**	[0.42,0.59]	0.50(0.03)**	[0.42,0.59]	0.50(0.03)**	[0.42,0.59]
Engage2 → Engage3	0.81(0.06)**	[0.65,0.97]	0.81(0.07)**	[0.64,0.98]	0.81(0.06)**	[0.65,0.97]
Achieve1→Engage1	0.40(0.04)**	[0.30,0.49]	0.40(0.04)**	[0.30,0.49]	0.40(0.04)**	[0.30,0.49]
Achieve3→Engage2	0.21(0.03)**	[0.12,0.29]	0.21(0.03)**	[0.13,0.28]	0.21(0.03)**	[0.12,0.29]
Achieve5→Engage3	0.13(0.05)**	[0.02,0.25]	0.14(0.06)*	[0.00,0.18]	0.13(0.05)**	[0.02,0.25]
SFA→Engage1	0.12(0.05)*	[-0.01,0.25]	0.12(0.05)*	[-0.01,0.25]	0.12(0.05)*	[-0.01,0.25]
SFA→Engage2	0.00(0.04)	[-0.10,0.10]	0.00(0.04)	[-0.10,0.10]	0.00(0.04)	[-0.10,0.10]
SFA→Engage3	0.81(0.06)	[-0.06,0.16]	0.05(0.04)	[-0.06,0.16]	0.05(0.04)	[-0.06,0.16]
Engage1→Achieve3	0.15(0.04)**	[0.04,0.25]	0.15(0.04)**	[0.05,0.25]	0.16(0.04)**	[0.05,0.26]
Engage2→Achieve5	0.03(0.03)	[-0.06,0.12]	0.03(0.04)	[-0.06,0.12]	0.03(0.04)	[-0.06,0.12]
Model fit modifications						
SFA→Achieve4	0.03(0.04)	[-0.07,0.13]	0.03(0.04)	[-0.07,0.13]		
SFA→Achieve5					0.00(0.04)	[-0.10,0.09]
Achieve4→Engage3			-0.01(0.06)	[-0.16,0.15]		
Indirect effects						
SFA→Achieve4→Achieve5→Engage3	0.01(0.01)	[-0.01,0.03]	0.01(0.01)	[-0.01,0.01]		
SFA→Achieve4→Engage3			0.01(0.01)	[-0.01,0.02]		
SFA→Achieve5→Engage3					0.00(0.01)	[-0.01,0.01]

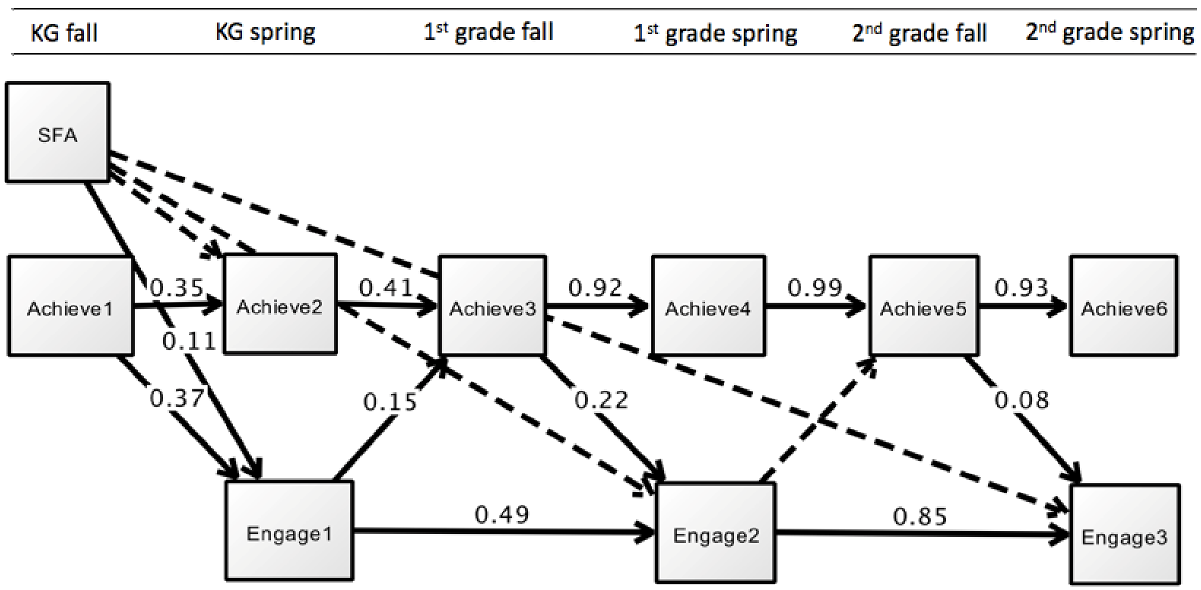

Figure 4. Completely standardized parameter estimates from the initial model (Model 1). Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, Black, and propensity strata 3.

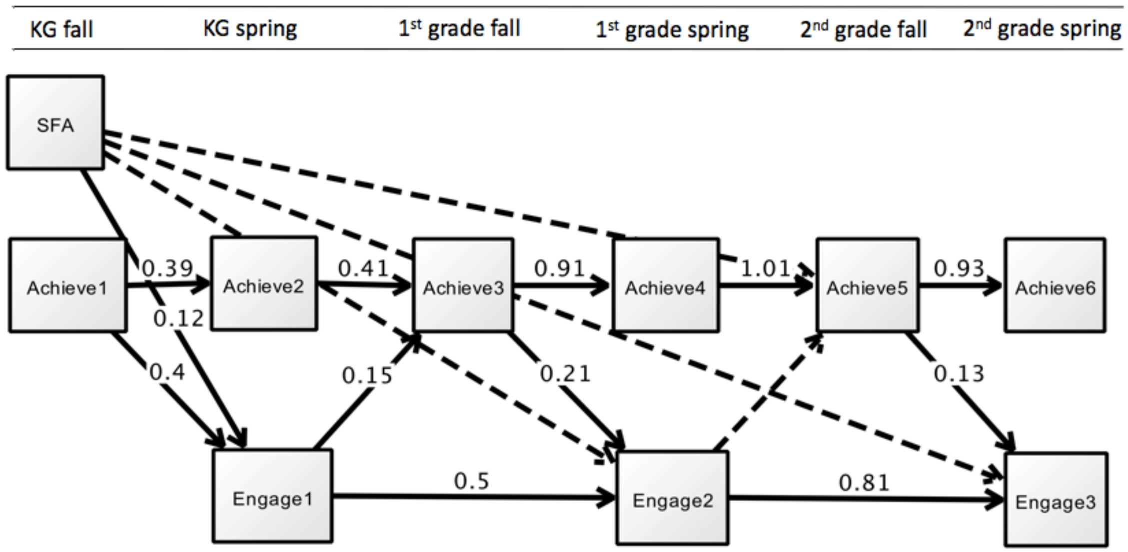

Figure 5. Completely standardized parameter estimates from best fitting model (Model 7). Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, Black, and propensity strata 3.

Chapter 5. Discussion

This study sought to investigate the role of Success For All in the development of students' non-cognitive factors. While the literature has explored academic effects of Success For All, the whole school components of the program give reason to believe that it may also have an impact on non-cognitive outcomes. Moreover, the literature linking non-cognitive outcomes and achievement predominantly assume that the former affects the latter or treat them as separate outcomes. However, other literature also indicates the reverse is likely true, in that achievement could affect non-cognitive factors. As a result, this study hypothesized that Success For All may influence students' non-cognitive factors through improved achievement.

While mediation is done widely in a cross-sectional manner, the rich longitudinal data that was available enabled the use of autoregressive cross-lagged models to map relationships between the mediators and outcomes over time.

Finding 1: SFA Effects on Student Engagement

The first finding in this study was that SFA had a small but positive relationship with teacher-reported engagement at the end of kindergarten. This was consistent across the cross-sectional, longitudinal, and ARCL models. The results partially confirm hypothesis 1. In summarizing the research on social and emotional learning programs, Greenberg et al. (2003) writes that "when classroom instruction is combined with efforts to create environmental support and reinforcement from peers, family members, school personnel, health professionals, other concerned community members, and the media, there is an increased likelihood that students will adopt positive social and health practices" (p. 470). Indeed, SFA's multiyear, multicomponent design that integrates efforts across multiple stakeholders in the school fits Greenberg et al.'s (2003) observation of programs that have shown success with SEL outcomes.

The longitudinal multilevel models did not confirm hypothesis 2. There was no significant interaction effect between SFA and time suggesting that there was no statistically significant difference between SFA and control students in the the rate of growth in any of the non-cognitive outcomes.

Possible explanations for relationship. While it is not the aim or design of this study to pinpoint which components of the program lead to student engagement in particular, the engagement literature leads to some conjectures.

Classroom environment. One possibility is that the strong professional development component that emphasizes proactive classroom management and a scripted curriculum provide an environment in which students can be more engaged. The literature suggests that the organizational climate of schools influences both students' engagement and their academic achievement (Dotterer & Lowe, 2011). Downer, Rimm-Kaufman, and Pianta (2007) explain: "varying classroom conditions impose different levels of behavioral and academic demands on children, a particularly relevant issue as the degree to which the classroom environment supports or detracts from behavioral engagement is considered" (p. 415). Classrooms characterized as demonstrating productive use of instructional time and low classroom chaos were linked to children being more engaged in classroom activities (NICHD Early Child Care Research Network, 2005). Similarly, in a study of 1,018 elementary students, students' reports of chaos in the social context of the classroom was negatively correlated with their reports of engagement; conversely, reports of structure in the classroom was positively correlated with reported engagement (Skinner, Kindermann, & Furrer, 2009).

Because SFA emphasizes a structured curriculum and professional development around strong classroom management skills, this may have contributed to a more orderly and thoughtful use of class time, minimizing opportunity for students to be distracted or to disengage from classroom activity. Part of the instructional process of SFA includes frequent checks for understanding (Correnti & Rowan, 2007), again reinforcing the possibility that minimal downtime in instruction may keep students more engaged. SFA teachers may have received more professional development and support from coaches around these techniques, which may have accounted for more engaged classrooms.

Cooperative learning. Furthermore, a distinctive instructional component of the SFA program is cooperative learning. Cooperative learning is embedded in instruction that engages students in rich discussion with one another about the reading and writing they are doing. Cooperative learning draws on social interdependence theory (Johnson & Johnson, 2009), which assumes that outcomes are influenced by individuals' own and others' actions. The research on cooperative learning has indicated that students tend to spend more time on task and demonstrate higher intrinsic motivation in cooperative learning situations than in competitive or individualistic situations (Johnson & Johnson, 1989). This is because successful cooperative learning implies that there is positive interdependence among students who are learning together such that students mutually encourage and help each other reach shared goals (Hermann, 2013). Empirical studies have also found a positive link between cooperative learning and student engagement (Hermann, 2013; Zhao & Kuh, 2004). In describing effective social and emotional programs, a report by the Aspen Institute (Aspen Institute, 2018) suggests that it is important to foster an environment in which competencies needed to interact socially are learned. The cooperative learning encouraged in SFA may allow for these competencies to develop and thus lead to greater engagement.

Different results for teacher-reported engagement and student-reported engagement. An interesting observation from the results was that a small positive relationship between SFA and teacher-reported engagement was found, but there were no effects on student-reported engagement. Why could there be differing relationships if both the teacher and student surveys are measuring a similar underlying construct? Part of the reason for this discrepancy may lie in the reliability of one mode over the the other. Research on student engagement measures has found that teacher-reported engagement measures demonstrate higher reliability than student-reported ones (Fredricks et al. 2011). For instance, two common measures of engagement, the Engagement versus Disengagement with Learning (EvsD) and Research Assessment Package for Schools (RAPS) reflect this trend. The EvsD reported cross-year correlations of .53–.68 for the student self-report measure and .65–.82 for the teacher report. The RAPS reports reliabilities for the student self-report ranging between .66 and .78 and a reliability of .87 for the teacher report. Indeed, differences in reliability were consistent with the literature in this study sample as well: teacher-reported engagement had a reliability coefficient of .96 while student-reported engagement was at a markedly lower 0.79.

In a validation study of the EvsD, Skinner, Kindermann, and Furrer (2009) found that correlations between children's observed behavior as rated by third-party observers and teacher-reported engagement were stronger than student reports. This suggests that teacher may be a more reliable source for assessing the engagement of students. This may be particularly true with younger students as in this sample, who may be developing their self-awareness.

In addition to possible measurement issues, it is also possible that SFA teachers were better prepared to evaluate children's engagement over time. Perhaps with the training they received from SFA they were able to notice more drastic changes in students' engagement in class. It is important to note that it is unlikely that SFA teachers would feel biased to rate their students higher on engagement compared to comparison students. This is because the original evaluation was led by a third-party unrelated to SFA and schools had elected to adopt SFA even before the inception of the study and thus were not persuaded by researchers to participate in SFA. Because of these reasons, it is unlikely that the difference found here in teacher-reported engagement is due to SFA teachers purposely rating their students higher to please the evaluators. Rather, it is more likely that the strong professional development teachers received may have helped them to better observe and recognize changes in student engagement, perhaps better than students were able to notice in themselves.

The relationship with teacher-reported engagement found only at kindergarten. The results suggest that there was a difference between SFA and control schools only at the end of kindergarten and not in other grades. While promising, the findings that there was a difference only at kindergarten is not ideal given the research that shows that student engagement tends to decline with each grade level (Gallup, 2016). The ideal scenario would involve students experiencing an increase in engagement compared to the control group in the first year and continuing this trajectory into the older grades as well.

One explanation for this phenomenon may be that behaviors are more malleable in younger than in older children (Fisak, Richard, & Mann, 2011; Lock & Barrett, 2003). That is, students may have been more impressionable to the different activities that made them more likely to react positively to the program. Another explanation may be that because kindergarten was the first year of the program, the novelty of the activities may have been more impactful, and gave a boost to students' engagement. Indeed, introducing a novel aspect or programmatic feature has been linked to heightened short term effects (Harrington, Hoyle, Feene, & Yungbluth, 2001).

Absence of effects on other outcomes. It is also important to note the absence of effects found with other outcomes. Although a significant difference in engagement (favoring SFA students) was found, no such differences were found with teacher-reported engagement, student-reported self-efficacy or teacher-reported antisocial behavior. A possible reason for the absence of effects with student-reported engagement was explored previously. However, it was unexpected that there were no program effects on the other non-cognitive factors (i.e., self-efficacy and anti-social behavior).

For student-reported self-efficacy, a similar limitation with student-reported engagement may have been possible in that the reliability of this scale was the lowest among all scales ($\propto$ = 0.60). This greater instability as a scale may account for lack of differences found between SFA and control students. Another possibility is that students generally professed high levels of self-efficacy. In fact, self-efficacy had the largest mean value among the outcome variables, with treatment students scoring an average of 3.26 out of 4 and control students scoring an average of 3.30 out of 4 on this scale in the first year. The overall high ratings of self-efficacy may have allowed little room for drastic differences between treatment and control groups.

In terms of teacher-reported antisocial behavior, it is unclear what discipline policies or school culture initiatives control schools implemented during the study. It is possible that control schools also took measures to reduce antisocial behavior through various interventions or initiatives, which may have accounted for the lack of differences found with SFA schools. It is also possible that in implementing the program, SFA teachers and staff focused more on the curricular aspects of the program rather than the social-emotional components such as the "Getting Along Together" curriculum. Again, without fidelity measures for specific program components, it is difficult to ascertain whether certain aspects were more emphasized than others, which may partly explain the lack of program effects on student behavior.

Finding 2: Relationship Between Non-Cognitive Outcomes and Achievement

Another area of interest that guided this study was clarifying the relationship between non-cognitive outcomes and achievement. As described in Chapter 2, the vast majority of studies treat non-cognitive outcomes as a predictor of academic achievement. A number of meta-analyses of SEL programs elucidate their effects on student achievement (Durlak et al., 2011; Corcoran, Cheung, Kim, & Xie, 2017; Payton et al., 2008). In these reviews, the assumption is that programming around non-cognitive factors and SEL is responsible for increased academic achievement (Zins, Bloodworth, Weissberg, & Walberg, 2004). While this is an empirically supported assumption, the reverse direction was hypothesized to be likely true based on theory and prior research (Miles & Stipek, 2006; Morgan, Farkas, Tufis, & Sperling, 2008). However, there is a scarcity in the literature that model both possibilities regarding directionality over time.

In this study, the use of an ARCL model allowed for such modeling. A few findings arose from the analysis around the relationship between non-cognitive outcomes and achievement over time.

Stability of variables over time. Consistent with the findings of Cunha and Heckman (2007), this study found evidence for high degrees of self-productivity for both reading achievement and teacher-reported engagement. Cunha and Heckman (2007) describe self-productivity as the phenomenon in which "higher stocks of skills in one period create higher stocks of skills in the next period" (p.10). That is, prior achievement strongly predicts later achievement and prior non-cognitive factors strong predict later non-cognitive outcomes. This stability of variables for reading achievement and teacher-reported engagement over time was expected and is consistent with the literature. This stability in variables was found also in ARCL models of the other non-cognitive outcomes as well (see Appendix C).

Interestingly, the relationship between later measurements of each variable were stronger than earlier measurements. For instance, the effect size between achievement in the fall of kindergarten to the end of kindergarten was +0.35, while the effect size between achievement in the fall of second grade to the end of second grade was a larger +0.93. A similar pattern was observed with non-cognitive outcomes as well; the effect size was +0.49 in the earlier measurements compared to +0.85 in the later measurements. A high stability coefficient, or correlation between one measurement in time and the next signifies that the change in individual differences is relatively small. This was true as students became older, suggesting that there were smaller changes within individuals with the passage of time. This also affirms the observation in the previous section that behaviors may be more malleable in earlier years. Similarly, Cunha and Heckman (2007) assert that returns to later investments in non-cognitive and cognitive skills are greater if higher early investment is made since behaviors and attitudes are more malleable at a younger age.

Significance of cross-lagged components. The second observation was that, as theory predicts, there were significant cross-lagged relationships. Both the longitudinal multilevel models and ARCL models confirmed that prior achievement predicted later engagement. This confirms hypotheses 3 and 4. The ARCL models further indicated that prior engagement predicted later achievement. Scholars have suggested a dynamic development of non-cognitive skills and achievement (Liu, 2016; Cunha & Heckman, 2006) whereby both variables influence one another over time. The engagement literature in particular has explored the effect of achievement on engagement as well as the reverse, finding a strong positive relationship between engagement and performance across diverse populations (Finn, 1989; Finn & Rock, 1997). These cross-lagged relationships were true also of the other non-cognitive outcomes (see Appendix C), though the relationships differed in magnitude and were significant at different time points.

Of particular interest here is that comparatively, the effect size of prior reading achievement on later teacher-reported engagement was larger than prior engagement on later achievement. In fact, there was no significant relationship found between engagement measured in first grade and achievement in second grade. The stronger relationship found between prior achievement and engagement was inconsistent with some extant literature. For instance, Hughes, Luo, Kwok, and Lloyd (2008) sought to model the relationship between engagement, reading and math achievement, and teacher-student relationship quality. Their model suggested that prior engagement had a larger effect on later math and reading achievement than vice versa. For instance, the first measurement of engagement had a standardized estimate of 0.09 on later reading while the first measurement of reading had a 0.04 estimate on later engagement. One explanation for this differing finding may be that the measure for engagement that the authors used was adopted from the Conscientiousness scale of the Big Five Inventory, which is more customarily used for personality assessments. The different conceptualization of engagement used in this study may account for these observed differences.

Despite the finding that the direction from reading achievement to later engagement was stronger in this study, it is important to note that later engagement was still strongly predicted by previous engagement. In fact, engagement at any time point was more strongly predicted by a previous measurement of engagement than by previous achievement. This is consistent with Liu (2016), who also found similar patterns between non-cognitive skills and achievement among young children.

In sum, the significance of cross-lagged relationships lends support, in terms of teacher-reported engagement, for the theory around the dynamic development of non-cognitive factors and achievement. The presence of cross-lagged relationships with the other non-cognitive factors (i.e., self-reported engagement, teacher-reported anti-social behavior, and student-reported self-efficacy) further corroborates this theory. The ARCL models also highlight the importance of considering *both* previous measurement of non-cognitive factors and achievement when observing later non-cognitive outcomes. The strength of this study's analytic design is that while cross-lagged relationships have been theorized, few empirical studies have sought to map out the relationships over time.

Finding 3: Absence of Mediation Between Achievement and Engagement

The third overarching finding with the ARCL models was that no mediation was found between reading achievement and teacher-reported engagement. To ensure that mediation relationships were comprehensively examined, the ARCL models looked at a combination of different possible direct effects from treatment to achievement and relationships across different time lags. These additional analyses did not find a significant mediating effect.

Interestingly, when reverse mediation was examined as a supplementary analysis (as non-cognitive factors were the focus outcomes of this study), the pathway from treatment affecting first grade fall achievement through teacher-reported engagement at kindergarten reached marginal significance ($\beta = 0.02$, $p < 0.10$). Although marginally significant, due to the relatively small sample size in this study this relationship is still worth considering. While not statistically significant, the results may warrant further investigation in the future that rigorously examines non-cognitive factors' mediating effect on achievement. For this sample, the hypothesized mediation path through achievement was not confirmed.

Indeed, the body of literature often treats achievement and non-cognitive outcomes as separate outcomes that theoretically impact one another. Increased SEL programming is thought to influence academic outcomes (Garcia, 2016; Payton et al., 2008) and academic achievement is thought to shape non-cognitive outcomes (Finn & Cox, 1992; Voelkl, 1997). The mediating effect was hypothesized, but this study adds further support to the notion that achievement and non-cognitive factors may be distinct outcomes that impact one another over time, but perhaps not *through* one another.

One puzzling observation from the ARCL models may be the absence of significant direct effects from SFA to achievement. This would contradict much of the literature (Borman et al., 2007; Madden, Slavin, Karweit, Dolan, & Wasik, 1993; Quint et al., 2015) that supports SFA's effects on student achievement. While a significant direct effect from SFA to engagement ($X \rightarrow Y$ relationship) was found and cross-lagged relationships between engagement and achievement found ($M \rightarrow Y$), the missing link from SFA to achievement ($X \rightarrow M$) may have accounted for the lack of significant mediation effects observed.

An important methodological difference between this study and other studies that have found SFA's academic effects is that the ARCL models account for every previous measurement of the mediating variable. Traditional multilevel models, such as the ones used in the Cheung and Slavin (2017) analysis, control for the initial pretest and consider the outcome as the difference in the final measurement accounting for the initial pretest. However, in the ARCL models explored in this study, each measurement of achievement was controlled for in order to observe relationships longitudinally. Because of this, the differences in achievement between SFA and control schools may have been rendered minimal. Put another way, if each previous measurement of the outcomes variable is accounted for, there is likely to be less of difference found between measurements than the difference found between simply the initial and final measurements.

As supplementary analysis to examine this conjecture, traditional multilevel models based on the cross-sectional multilevel models explored earlier revealed that simply controlling for kindergarten fall achievement scores, SFA schools did have an effect on 2^{nd} grade fall achievement. In fact, the effect size was ES=+0.23 (b =9.96, p<.05), similar to the effect size of +0.26 found by Cheung and Slavin (2017) in the second year. The results of this analysis are not expected to match their analysis exactly: the authors used a different method for defining a control group, used a different propensity score method (one-to-one matching), and had slightly different analytic models. Mediation with this cross-sectional model was run as supplementary analysis, but here too, achievement did not significantly mediate the relationship between SFA and engagement (b=0.01, p>.05). This lends more credence to the main findings that no mediation was found.

Why was the ARCL, or SEM, approach used in this study instead of traditional multilevel models to examine mediation? It is worth reviewing the decision to use ARCL models instead of another approach to explore mediation. Longitudinal SEM enables analysis of relationships within waves of data and changes across waves of data simultaneously (MacKinnon, 2008). In a similar vein, Selig and Little (2012) explain that ARCL models are advantageous in modeling potentially reciprocal relationships and specifying hypothesized directions of effects over time. It is particularly useful in this context in which modeling the relationship between achievement and non-cognitive outcomes at several different time points was of interest. Because of this advantage, a longitudinal SEM approach to mediation was preferred to the traditional multilevel approach.

One may wonder why multilevel structural equation modeling (MSEM) was not used instead of ARCL models. MSEM is gaining popularity as an analytic method; however, in a recent article, McNeish (2017) noted that while MSEM offers greater flexibility in modeling than traditional multilevel models, the vast majority of studies that utilize this method often overlook the sample size requirements (100 clusters). As there is discussion in the current literature around sample size requirements, MSEM may not have been the most appropriate model for the current sample.

Thus, there were characteristics of the design, purpose, and sample of the study that justified the use of ARCL mediation models even though it departs from the traditional multilevel analytic approach.

Limitations

Several limitations of the study should be noted. One limitation is that even though the study used propensity score methods to account for the quasi-experimental design, it is not possible to get to perfect causal claims. The lack of random assignment into treatment and control groups leads to less definitive conclusions about causality than do randomized control trials (Shadish, Cook, & Campbell, 2002), which are though to be a more robust method for examining causality.

Another limitation is that implementation measures were not systematically collected in the larger study. While this study suggests that SFA students were more engaged than comparison students at the end of kindergarten, it is difficult to identify which aspects of the program may have contributed to this. The original study from which the current study has been derived collected and instructional logs to examine teachers' literacy practices, but it did not collect data on activities outside the classroom. The SFA model involves much intentional activity outside of isolated teacher instruction, so these logs would not capture the full extent of SFA in schools. Qualitative interviews of students, staff, or other stakeholders may also have better indicated aspects of SFA that were particularly engaging.

Moreover, the study's sample consisted of predominantly disadvantaged students. While this study was designed purposefully to be representative of most urban education settings, it is not meant to be national representative. The current study could be replicated with a more diverse sample to make more general claims about the overall national population.

Relatedly, the measures used for this study have not been used previously in other studies. The psychometrics of the measures have not been tested in other studies. However, the reliabilities were for the most part strong in this sample (with the exception of reading self-efficacy) and the questions generally demonstrated face validity. Further studies validating this instrument would add confidence in the results of this study.

Implications for practice

Despite the limitations, this study contributes to the literature and future practice in two broad ways. First, it suggests the promise of programs that are not explicitly SEL-focused in improving students' non-cognitive outcomes and engagement in particular. SFA is not packaged or widely known as an SEL program; instead, it takes a whole-school approach to change not only instructional practices in the classroom but also shift the school culture as a way of improving student outcomes holistically. While SFA is known more as an academic program, its program components may have a wider reach that touches on non-cognitive factors as well. As the concept of SEL and non-cognitive factors continues to gain prominence in the education field, schools and local education agencies will need to respond by adopting practices or programs that will address students' non-cognitive factors. Instead of starting and testing (often costly) stand-alone SEL programs in response to interest in SEL, perhaps schools can examine what resources their schools are currently using and leverage instruction changing and school-culture shifting approaches. The Aspen Institute (2018) also advocates for this approach as it suggests that "a lens of integration that encompasses the entire school community are likely to be the most effective and sustainable and are less likely to be considered an add-on or nice-to-have" (p. 11). The study also echoes the theme in literature emphasizing the need to intervene especially in earlier years when students' achievement and non-cognitive factors may be malleable, and make efforts to continue this growth.

Second, this study helps to clarify the relationship between achievement and teacher-reported engagement over time. Zins et al. (2004) writes that “the promotion of social emotional learning goals is no longer seen as ‘separate’ or even parallel to the academic mission of schools” (p. 9). This suggests that the field increasingly views achievement and social emotional learning, or non-cognitive factors, as inextricably tied. This is confirmed by the current study, where cross-lagged relationships were observed over time. Thus, instead of a one-sided view in which there are efforts to improve achievement and separate efforts to improve non-cognitive factors, the study and the broader literature encourage efforts to integrate both. Indeed, the Aspen Institute (2018) posits that social and emotional learning and academic achievement should not be seen as two separate goals. They argue that “decades of research in human development, cognitive and behavioral neuroscience, and educational practice and policy have illuminated that major domains of human development—social, emotional, cognitive, linguistic, academic—are deeply intertwined in the brain and in behavior” (p. 10). In this sense, there is a mismatch in practice and research: in practice, local education agencies and states are moving more towards integrating the two, while in research they are still oftentimes reported and treated as separate outcomes.

By adopting a more integrated approach, research in this field may be able to more effectively consider ways to foster a holistic education for students.

Future directions

Currently, there is general agreement around the importance of educating the whole child. This section proposes a few directions for future research considering the current questions and interests in the field. One next step in the field of research in non-cognitive factors includes connecting the promotion of non-cognitive skills with equity. That is, are current programs and approaches supporting the development of non-cognitive skills for all students? The literature is clear that students from disadvantaged backgrounds come in with poorer academic and non-cognitive skills (Duncan & Murnane, 2011; Garcia, 2015). Indeed, even in this sample there were discrepancies in non-cognitive outcomes by racial and socioeconomic status, favoring non-minority and higher SES students. As the diversity of students' racial and socioeconomic backgrounds continues to increase in the U.S. (U.S. Department of Education, 2017), an important priority in research and practice in educating the whole child will have to link program effects with equitable outcomes across all students.

Another direction for this line of research involves teacher preparation and ongoing professional development. In addition to the importance of curricular and school culture supports by programs such as SFA, the research suggests that teachers' own non-cognitive skills are important factors in students' non-cognitive development. For instance, Jennings and Greenberg (2009) put forth a theoretical model that asserts that teachers' social and emotional competence is pivotal for social and emotional learning (SEL) program effectiveness. Next steps in the research may include examining whether programs such as SFA or others designed to improve student non-cognitive growth either explicitly or implicitly affect teachers' own non-cognitive factors. Such programs are likely to be more effective when teachers can model the same non-cognitive factors that they hope to espouse in their students. As the Aspen Institute Report (2017) describes: "It's hard for someone to give what they don't have. You can't assume that, just because they're adults, they have the skills and the mindsets they need to model healthy behaviors and understand the core knowledge of social-emotional learning. It's a wonderful thing when adults and kids can grow together" (p. 12). Future evaluations of whole school approaches or programs that nurture the whole child should be mindful of teachers' own non-cognitive competencies and how they help or hinder students' growth.

In a similar vein, while a mediating relationship between achievement on SFA and non-cognitive outcomes was not found, perhaps it is accomplished through other mediators such as an improved school climate, which has also been found in the research to mediate such outcomes (Liu, Van Damme, Gielen, & Van Den Noortgate, 2015; Uline & Tschannen-Moran, 2008). Unfortunately, for this study there were no theoretically or empirically reliable measures of school culture, and this precluded testing of this possibility. However, future research may explore other possible mediators or explanations for SFA's effect on non-cognitive skills.

The current study identifies the potential of the SFA model in promoting teacher-reported engagement, an important non-cognitive factor, at the end of kindergarten. While these results reveal novel insights in the field of non-cognitive programming, they further open a wealth of questions and new directions that will likely advance the field as a whole.

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education*, *70*(2), 87-107.

Anderson, C. S. (1982). The search for school climate: A review of the research. *Review of Educational Research*, *52*(3), 368-420.

Astor R. A., Guerra N., Van Acker R. (2010). How can we improve school safety research? *Educational Researcher, 39*, 69–78.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399-424.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359-373.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117-148f.

Battistich, V., Schaps, E., & Wilson, N. (2004). Effects of an elementary school intervention on students' "connectedness" to school and social adjustment during middle school. *The Journal of Primary Prevention*, *24*(3), 243-262.

Bavarian, N., Lewis, K. M., DuBois, D. L., Acock, A., Vuchinich, S., Silverthorn, N., Snyder, F.J., Day, J., Ji, P. & Flay, B. R. (2013). Using social-emotional and character development to improve academic outcomes: A matched-pair, cluster-randomized controlled trial in low-income, urban schools. *Journal of School Health*, *83*(11), 771-779.

Becker, B. E., & Luthar, S. S. (2007). Peer-perceived admiration and social preference: Contextual correlates of positive peer regard among suburban and urban adolescents. *Journal of Research on Adolescence*, *17*(1), 117-144.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257-278.

Becker, G. S. (1964). *A theoretical and empirical analysis, with special reference to Education.* New York, NY: Columbia University Press.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal*, *2*(4), 358-377.

Beets, M. W., Flay, B. R., Vuchinich, S., Snyder, F. J., Acock, A., Li, K. K., & Durlak, J. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. *American Journal of Public Health*, *99*(8), 1438-1445.

Bentley, J. P. (2011). *An examination of statistical methods for longitudinal mediation modeling*. The University of Alabama at Birmingham. Retrieved from http://search.proquest.com/openview/0a218171ab0097f5209dbab79efb9e6d/1?pq-origsite=gscholar&cbl=18750&diss=y

Berkowitz, M. W., & Bier, M. C. (2007). What works in character education. *Journal of Research in Character Education*, *5*(1), 29.

Bidwell, C. E., & Yasumoto, J. Y. (1999). The collegial focus: Teaching fields, collegial relationships, and instructional practice in American high schools. *Sociology of Education*, *72*(4), 234-256.

Bierman, K. L., Coie, J., Dodge, K., Greenberg, M., Lochman, J., McMohan, R., & Pinderhughes, E. (2013). School outcomes of aggressive-disruptive children: prediction from kindergarten risk factors and impact of the fast track prevention program. *Aggressive Behavior*, *39*(2), 114-130.

Binet, A., & Simon, T. (1916). *The development of intelligence in children: The Binet-Simon scale.* Baltimore, MD: Williams & Wilkins Company.

Bloom, B. (1976). *Human characteristics and school learning.* New York, NY: McGraw-Hill.

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of Educational Psychology*, *90*(1), 102.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972-1059.

Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost-effectiveness of Success For All. *Educational Evaluation and policy analysis*, *24*(4), 243-266.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125-230.

Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success For All. *American Educational Research Journal*, *44*(3), 701-731.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America* (Vol. 57). New York: Basic Books.

Bradshaw, C.P., Mitchell, M.M., & Leaf, P.J. (2010). Examining the effects of School-Wide Positive Behavioral Interventions and Supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12* (3), 133-148.

Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), *Theorectical models of human development* (5 ed., pp. 993-1028). (Handbook of Child Psychology; Vol. 1). New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-136.

Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.

Butler, J. K. (1999). Trust expectations, information sharing, climate of trust, and negotiation effectiveness and efficiency. *Group and Organization Management*, *24*(2), 217-238.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. New York, NY: Russell Sage Foundation.

Cacioppo, J. T., Hawkley, L. C., & Thisted, R. A. (2010). Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychology And Aging, 25(2), 453-463. doi:10.1037/a0017216

Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce.* Retrieved from http://files.eric.ed.gov/fulltext/ED519465.pdf.

Castrechini, S., & London, R. A. (2012). *Positive student outcomes in community schools.* Retrieved from https://www.americanprogress.org/issues/education/reports/2012/02/22/11098/positive-student-outcomes-in-community-schools/.

Chamberlain, A., Daniels, C., Madden, N., & Slavin, R. (2007). A randomized evaluation of the Success For All middle school reading program. *Middle Grades Research Journal, 2*(1), 1-21.

Chen, X., Huang, X., Chang, L., Wang, L., & Li, D. (2010). Aggression, social competence, and academic achievement in Chinese children: A 5-year longitudinal study. *Development and Psychopathology*, *22*(3), 583-592.

Chen, X., Rubin, K. H., & Li, D. (1997). Relation between academic achievement and social adjustment: Evidence from Chinese children. *Developmental Psychology*, *33*(3), 518.

Cheung, A. C., & Slavin, R. E. (2016). Effects of Success For All on reading achievement: A secondary analysis using data from the Study of Instructional Improvement (SII). *AERA Open*, *2*(4), 1-10.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*, 295-313

Coie, J. D., & Jacobs, M. R. (1993). The role of social context in the prevention of conduct disorder. *Development and Psychopathology*, *5*(1-2), 263-275.

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*(5), 648.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Cook, T. D., Habib, F. N., Phillips, M., Settersten, R. A., Shagle, S. C., & Degirmencioglu, S. M. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, *36*(3), 543-597.

Cook, T. D., Murphy, R. F., & Hunt, H. D. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*(2), 535-597.

Corcoran, R. P., Cheung, A., Kim, E., & Xie, C. (2017). Effective Universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*.

Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, *48*(1), 236-264.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis*, *29*(4), 262-295.

Correnti, R., & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. American educational research journal, 44(2), 298-339.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31-47.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529–569.

Dalton, B. W. (2010). Antisocial and prosocial behavior. In Rosen et al. (Eds), *Noncognitive skills in the classroom: New Perspectives on Educational Research* (pp. 145-168). Research Triangle Park, NC: RTI International.

Di Giunta, L., Alessandri, G., Gerbino, M., Kanacri, P. L., Zuffiano, A., & Caprara, G. V. (2013). The determinants of scholastic achievement: The contribution of personality traits, self-esteem, and academic self-efficacy. *Learning and individual Differences*, *27*, 102-108.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 222.

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. *Journal of youth and adolescence*, *40*(12), 1649-1660.

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children's risk for school problems contribute to children's behavioral engagement in learning?. *School Psychology Review*, *36*(3), 413.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087.

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, *104*(2), 439.

Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In Duncan & Murnane (Eds), *Whither opportunity* (pp. 47-70). New York, NY: Russell Sage Foundation.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity?: Rising inequality, schools, and children's life chances*. New York, NY: Russell Sage Foundation.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Dusenbury, L., Calin, S., Domitrovich, C., & Weissberg, R. P. (2015). *What does evidence-based instruction in social and emotional learning actually look like in practice?* Retreived from https://static1.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/56374ac1e4b05d222e9b4dea/1446464193894/CASEL+Brief--What+Does+SEL+Look+Like+in+Practice--11-1-15.pdf.

Eliot, M., Cornell, D., Gregory, A., & Fan, X. (2010). Supportive school climate and student willingness to seek help for bullying and threats of violence. *Journal of School Psychology*, *48*(6), 533-553.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*(1), 128-141.

Farrington, C.A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T.S., Johnson, D.W., & Beechum, N.O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological methods*, *14*(1), 43.

Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban high school*. Albany, NY: SUNY Press.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117-142.

Finn, J. D., & Cox, D. (1992). Participation and withdrawal among fourth-grade pupils. American Educational Research Journal, 29, 141–162.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology*, *82*(2), 221.

Fisak, B. J., Richard, D., & Mann, A. (2011). The prevention of child and adolescent anxiety: A meta-analytic review. *Prevention Science*, *12*(3), 255-268.

Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Health Behavior*, *27*(1), S6-S21.

Flay, B. R., Allred, C. G., & Ordway, N. (2001). Effects of the Positive Action program on achievement and discipline: Two matched-control comparisons. *Prevention Science*, *2*(2), 71-89.

Flay, B. R., DuBois, D. L., & Ji, P. (2007). *Progress report of the randomized trial of Positive Action in Chicago: End of third year of intervention (Grade 5, Spring 2006).* Retrieved from http://ies. ed. gov/pubsearch.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring student engagement in upper elementary through high school: A description of 21 instruments. Issues & Answers. REL 2011-No. 098. Regional Educational Laboratory Southeast.

Gallup. (2016). *2016 Gallup student poll snapshot report*. Retrieved from http://news.gallup.com/reports/210995/6.aspx?g_source=WWWV7HP&g_medium=topic&g_campaign=tiles

Garcia, E. (2015). Inequalities at the starting gate: Cognitive and noncognitive skills gaps between 2010-2011 kindergarten classmates. Report. *Economic Policy Institute*. Retrieved from http://eric.ed.gov/?id=ED560407

García, E. (2016). The need to address non-cognitive skills in the education policy agenda. In J. Khine, M.S & Areepattamannil, S. (Eds.) *Non-cognitive Skills and Factors in Educational Attainment* (pp. 31-64). Rotterdam, Netherlands: Sense Publishers.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 243-259).

Greenberg, M. T., Weissberg, R. P., O'brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American psychologist*, *58*(6-7), 466-474.

Gregory, A., & Cornell, D. (2009). “Tolerating” adolescent needs: Moving beyond zero tolerance policies in high school. *Theory Into Practice*, *48*(2), 106-113.

Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School climate and implementation of a preventive intervention. *American Journal of Community Psychology*, *40*(3-4), 250-260.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher*, *39*(1), 59-68.

Guo, P. (2012). *School culture: A validation study and exploration of its relationship with teachers' work environment* (Doctoral dissertation, Fordham University).

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, *3*(3), 231-256.

Gutman, L. M., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people.* Retrieved from https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Lit_Review_Non-CognitiveSkills.pdf

Gutman, L. M., Brown, J., Akerman, R., & Obolenskaya, P. (2010). *Change in wellbeing from childhood to adolescence: risk and resilience.* Retrieved from http://eprints.ioe.ac.uk/19040/1/WBLResRep34.pdf

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement, 9*(2), 157-191.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625-638.

Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J., & Yungbluth, S. C. (2001). Evaluation of the All Stars character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. *Health Education & Behavior*, *28*(5), 533-546.

Hausman, J. A., & Wise, D. A. (1979). Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, 455-473.

Haynes, N. M., Emmons, C., & Ben-Avie, M. (1997). School climate as a factor in student adjustment and achievement. *Journal of Educational and Psychological Consultation*, *8*(3), 321-329.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*(3), 289-324.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451-464.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review*, *91*(2), 145-149.

Heckman, J. J., & Stixrud, U.S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, *24*(3), 411–482.

Herrmann, K. J. (2013). The impact of cooperative learning on student engagement: Results from an intervention. *Active Learning in Higher Education*, *14*(3), 175-187.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11-30.

Hillygus, D. S., Holbein, J. B., & Snell, S. (2016). *The nitty gritty: The unexplored role of grit and perseverance in voter turnout.* Retrieved from https://ssrn.com/abstract=2675326

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127-155.

Holden, G., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: A meta-analysis. *Psychological Reports*, *66*(3), 1044-1046.

Holen, S., Waaktaar, T., Lervåg, A., & Ystgaard, M. (2013). Implementing a universal stress management program for young school children: Are there classroom climate or academic effects? *Scandinavian Journal of Educational Research*, *57*(4), 420-444.

Hong, S., Yoo, S. K., You, S., & Wu, C. C. (2010). The reciprocal relationship between parental involvement and mathematics achievement: Autoregressive cross-lagged modeling. *The Journal of Experimental Education*, *78*(4), 419-439.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133-144.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hughes, J. N., Luo, W., Kwok, O. M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of educational psychology*, *100*(1), 1-14.

Humphrey, N. (Ed.). (2013). *Social and emotional learning: A critical appraisal*. London: SAGE Publications Limited.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. Review of Educational Research, 79, 491–525.

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly*, *22*(3), 314.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN, US: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, *38*(5), 365-379.

Jones, E. M., Gottfredson, G. D., & Gottfredson, D. C. (1997). Success for some: An evaluation of a Success For All program. *Evaluation Review*, *21*(6), 643-670.

Jones, S. M., Brown, J. L., Hoglund, W. L., & Aber, J. L. (2010). A school-randomized clinical trial of an integrated social–emotional learning and literacy intervention: Impacts after 1 school year. *Journal of Consulting and Clinical Psychology*, *78*(6), 829.

Kee, A. N. (2012). Feelings of Preparedness Among Alternatively Certified Teachers: What Is the Role of Program Features? *Journal of Teacher Education*, *63*(1), 23–38.

Kirsch, I., De Jong, J., LaFontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2003). *Reading for change: Performance and engagement across countries: Results of PISA 2000.* Retrieved from https://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33690904.pdf

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, *74*(7), 262-273.

Kline, R.B. (2005), *Principles and Practice of Structural Equation Modeling (2nd Edition ed.).* New York: The Guilford Press.

Kutsyuruba, B., Klinger, D. A., & Hussain, A. (2015). Relationships among school climate, school safety, and student achievement and well-being: a review of the literature. *Review of Education*, *3*(2), 103-135.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal*, *36*(5), 807-830.

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology*, *101*(1), 190.

Lai, M. H., & Kwok, O. M. (2015). Examining the rule of thumb of not using multilevel modeling: The “design effect smaller than two” rule. *The Journal of Experimental Education*, *83*(3), 423-438.

Lane, K. L., & Wehby, J. H. (2002). Addressing antisocial behavior in the schools: A call for action. *Academic Exchange,* 4-9.

Lee, J., & Shute, V. J. (2009). *The influence on noncognitive domains on academic achievement in K-12.* Retrieved from http://files.eric.ed.gov/fulltext/ED507799.pdf

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 172-192.

Lennon, J. M. (2010). Self-efficacy. In Rosen et al. (Eds), *Noncognitive skills in the classroom: New Perspectives on Educational Research* (pp. 145-168). Research Triangle Park, NC: RTI International.

Levine, M., & Ensom, M. H. (2001). Post hoc power analysis: an idea whose time has passed? *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *21*(4), 405-409.

Li, Y., Zhang, W., Liu, J., Arbeit, M. R., Schwartz, S. J., Bowers, E. P., & Lerner, R. M. (2011). The role of school engagement in preventing adolescent delinquency and substance use: A survival analysis. *Journal of Adolescence*, *34*(6), 1181-1192.

Liu, A. (2016). *Non-cognitive skills and the growing achievement gap*. Retrieved from https://www.psc.isr.umich.edu/pubs/pdf/rr16-861.pdf

Liu, H., Van Damme, J., Gielen, S., & Van Den Noortgate, W. (2015). School processes mediate school compositional effects: model specification and estimation. *British Educational Research Journal*, *41*(3), 423-447.

Lock, S., & Barrett, P. M. (2003). A longitudinal study of developmental differences in universal preventive intervention for child anxiety. *Behaviour Change*, *20*(4), 183-199.

Lunenburg, F. C., & Ornstein, A. C. (2011). *Educational administration: Concepts and practices*. Boston, MA: Cengage Learning.

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: a longitudinal panel analysis. *Journal of Adolescence*, *27*(2), 165-179.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86-92.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor and Francis Group.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success For All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal*, *30*(1), 123-148.

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, *46*(5), 816-841.

McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic failure, and school climate: A critical review. *Journal of Emotional and Behavioral Disorders*, *8*(3), 130-140.

McNeely, C. A., Nonnemaker, J. M., & Blum, R. W. (2002). Promoting school connectedness: Evidence from the national longitudinal study of adolescent health. *Journal of School Health*, *72*(4), 138-146.

McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 609-625.

Mehan, H., Villanueva, I., Hubbard, L., Lintz, A., & Okamoto, D. (1996). *Constructing school success: The consequences of untracking low achieving students.* New York: Cambridge University Press.

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development, 77*(1), 103–117.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, *41*(5), 417-436.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30-38

Muñoz, M. A., & Dossett, D. H. (2004). Educating students placed at risk: Evaluating the impact of Success For All in urban settings. *Journal of Education for Students Placed at Risk*, *9*(3), 261-277.

Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review*, *20*(4), 311-320.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. Sociological Methodology, *25,* 267–316.

National School Climate Cetner (2017). *School climate*. Retrieved from http://www.schoolclimate.org/climate/

NICHD Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. The Elementary School Journal, 105, 305–323. doi:10.1086/428746.

Noddings, N. (2005). What does it mean to educate the whole child? *Educational Leadership*, *63*(1), 8.

O'Keefe, D. J. (2007). Brief report: post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures*, *1*(4), 291-299.

Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*(2), 224-243.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, *42*(2), 190.

Pardini, D. A., Loeber, R., & Stouthamer-Loeber, M. (2005). Developmental shifts in parent and peer influences on boys' beliefs about delinquent behavior. *Journal of Research on Adolescence*, *15*(3), 299–323.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29-48.

Payton, J., Weissberg, R.P., Durlak, J.A., Dymnicki, A.B., Taylor, R.D., Schellinger, K.B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Pitts, S. C., West, S. G., & Tein, J. Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning*, *19*(4), 333-350.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209.

Price, H. E. (2012). School principal-staff relationship effects on school climate. *Interpersonal Relationships in Education*, 103-118.

Quint, J., Balu, R., DeLaurentis, M., Rappaport, S., Smith, T.J., & Zhu, P. (2013). *The Success For All model of school reform: early findings from the Investing in Innovation (i3) scale-up.* Retrieved from http://www.mdrc.org/publication/success-all-model-school-reform

Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. (2015). *Scaling Up the Success For All Model of School Reform: Final Report from the Investing in Innovation (i3) Evaluation.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2667101

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, *23*(1), 10-26.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353.

Rivers, S. E., Brackett, M. A., Reyes, M. R., Elbertson, N. A., & Salovey, P. (2013). Improving the social and emotional climate of classrooms: A clustered randomized controlled trial testing the RULER approach. *Prevention Science*, *14*(1), 77-87.

Rosen, J. A., Glennie, E. J., Dalton, B. W., Lennon, J. M., & Bozick, R. N. (2010). *Noncognitive skills in the classroom: New perspectives on educational research*. Research Triangle Park, NC: RTI International.

Ross, S., & Smith, L. (1994). Effects of the Success For All model on kindergarten through second-grade reading achievement, teachers' adjustment, and classroom-school climate at an inner-city school. *The Elementary School Journal, 95*(2), 121-138.

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal*, *44*(2), 252-297.

Rowan, B., Miller, R., & Camburn, E. (2009). *School improvement by design: Lessons from a study of comprehensive school reform programs.* Retrieved from http://repository.upenn.edu/cpre_researchreports/54

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39* (6), 1020–1035.

Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education program on pre-and early adolescents' well-being and social and emotional competence. *Mindfulness*, *1*(3), 137-151.

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology,73*(1), 93–105.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*(2), 149-174.

Schunk, D. H. (2015). Self-efficacy: Educational aspects. In Smelser & Baltes (Eds), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 13820-13822). Amsterdam, Netherlands: Elsevier.

Schwarzer, R., & Fuchs, R. (1996). Self-efficacy and health behaviours. *Predicting health behavior: Research and Practice with Social Cognition Models*, 163-196.

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In Larsen, Little, & Card (Eds). *Handbook of Developmental Research Methods* (pp. 265-278). New York, NY: Guilford Press.

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2-3), 144-164.

Shouse, R., Schneider, B., & Plank, S. (1992). Teacher assessments of student effort: Effects of student characteristics and school type. *Educational Policy*, *6*(3), 266-288.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, *69*(3), 493-525.

Slavin, R. E. (2008). Comprehensive school reform. In Good, T.L. (Ed.) *21st century education: A reference handbook* (pp. 211-259). Thousand Oaks, CA: Sage.

Slavin, R. E., Madden, N. A., & Chambers, B. (2009). *One million children: Success For All*. Thousand Oaks, CA: Corwin Press.

Snyder, F., Vuchinich, S., Acock, A., Washburn, I., Beets, M., & Li, K. K. (2010). Impact of the Positive Action program on school-level indicators of academic achievement, absenteeism, and disciplinary outcomes: A matched-pair, cluster randomized, controlled trial. *Journal of Research on Educational Effectiveness*, *3*(1), 26.

Spence, M. (1974). *Market signalling.* Cambridge, MA: Harvard University Press.

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. *Oxford Bibliographies. doi*, *10*, 9780199756810-0108.

Stipek, D., & Miles, S. (2008). Effects of aggression on achievement: Does conflict with the teacher make it worse? *Child Development*, *79*(6), 1721-1735.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357-385.

The Aspen Institute. (2018). *How learning happens: Supporting students' social, emotional, and academic development*. Retrieved from https://www.aspeninstitute.org/publications/learning-happens-supporting-students-social-emotional-academic-development/

Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology*, *39*(3), 217-228.

U.S. Department of Education (2017). Racial/ethnic enrollment in public schools. Retrieved from https://nces.ed.gov/programs/coe/pdf/coe_cge.pdf

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration, 46*, 55–73.

Voelkl, K. E. (1997). Identification with school. *American Journal of Education*, 105, 294–318.

Wang, M. T., & Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child Development*, *85*(2), 722-737.

Wang, M. T., Selman, R. L., Dishion, T. J., & Stormshak, E. A. (2010). A tobit regression analysis of the covariation between middle school students' perceived school climate and behavioral problems. *Journal of Research on Adolescence*, *20*(2), 274-286.

Weissberg, R. P., Goren, P., Domitrovich, C., & Dusenbury, L. (2013). CASEL guide effective social and emotional learning programs: Preschool and elementary school edition. *Chicago, IL: CASEL*.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp.56-75). Washignton, DC: Sage.

What Works Clearinghouse (2017). *Success For All.* Retrieved from https://ies.ed.gov/ncee/wwc/InterventionReport/672

White, N. A., & Loeber, R. (2008). Bullying and special education as predictors of serious delinquency. *Journal of Research in Crime and Delinquency*, *45*(4), 380-397.

Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998). Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. *Academy of Management Review*, *23*(3), 513-530.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432-445.

Willms, J. D. (2003). *Student engagement at school: A sense of belonging and participation.* Retrieved from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33689437.pdf

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*(4), 695-719.

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in higher education*, *45*(2), 115-138.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82-91.

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.) (2004). *Building school success through social and emotional learning: What does the research say?* New York, NY: Teachers College Press

Appendix A

Table A1.

Teacher-reported engagement scale reliability and items

	Item
sr1a	This student... Is eager to learn
sr1b	This student... Usually pays attention in class
sr1d	This student... Completes school work in an organized way
sr1f	This student... Works well independently
sr1g	This student... Wants to do well in school
sr1h	This student... Keeps his/her personal belongings organized
sr1j	This student... Works hard on school assignments
sr1l	This student... Persists when work is difficult
sr1n	This student... Usually completes work on time
sr1o	This student... Uses free time in constructive ways
sr1q	This student... Works carefully and methodically
Reliability	0.96

Table A2.

Teacher-reported antisocial behavior scale reliability and items

	Item
sr1e	This student... Often talks back to adults
sr1m	This student... Gets angry easily
sr1c	This student... Frequently argues with others
sr1k	This student... Disrupts the work of others
sr1r	This student... Gets into fights with other children
sr1i	This student... Often acts impulsively
sr1p	This student... Sometimes damages property
Reliability	0.93

Table A3.

Student-reported engagement scale reliability and items

	Item
sm6pr	It's hard for me to finish my work in reading
sm8pr	It's hard for me to listen during reading class
sm14pr	It's hard for me to pay attention in reading class
sm2m	It's hard for me to listen during math class
sm11pm	It's hard for me to finish my work in math

sm17pm	It's hard for me to pay attention in math class
sm4pr	Get in trouble-talking/disturbing others - reading
sm20pm	Get in trouble-talking/disturbing others - math
Reliability	0.79

Table A4.

Student-reported reading self-efficacy scale reliability and items

	Item
sm1r	I'm good at reading
sm9r	I do well in reading
sm15r	Work in reading is easy for me
sm21r	I learn things quickly in reading
Reliability	0.60

Appendix B

Table B1.

Correlations among outcome variables at different time points.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Teacher reported engagement (kindergarten)	1.00											
2. Teacher reported engagement (1st grade)	0.60**	1.00										
3. Teacher reported engagement (2nd grade)	0.52**	0.56**	1.00									
4. Student reported engagement (kindergarten)	0.11	0.09	0.09	1.00								
5. Student reported engagement (1st grade)	0.18**	0.21**	0.29**	0.22**	1.00							
6. Student reported engagement (2nd grade)	0.28**	0.29**	0.28**	0.16*	0.25**	1.00						
7. Antisocial behavior (kindergarten)	0.55**	0.39**	0.35**	0.10	0.12	0.23**	1.00					
8. Antisocial behavior (1st grade)	0.44**	0.61**	0.42**	0.06	0.17**	0.28**	0.65**	1.00				
9. Antisocial behavior (2nd grade)	0.44**	0.51**	0.58**	0.10	0.21**	0.24**	0.55**	0.65**	1.00			
10. Self-efficacy (kindergarten)	0.12*	0.09	0.15	-0.03	0.12	0.06	0.02	0.00	0.02	1.00		
11. Self-efficacy (1st grade)	0.24**	0.20**	0.18*	0.03	0.15*	0.11	0.13	0.10	0.08	0.19**	1.00	
12. Self-efficacy (2nd grade)	0.11	0.15	0.16	0.03	0.08	0.21**	0.04	0.00	0.03	0.10	0.26**	1.00

Note. *p<.05 **p<.01

Appendix C

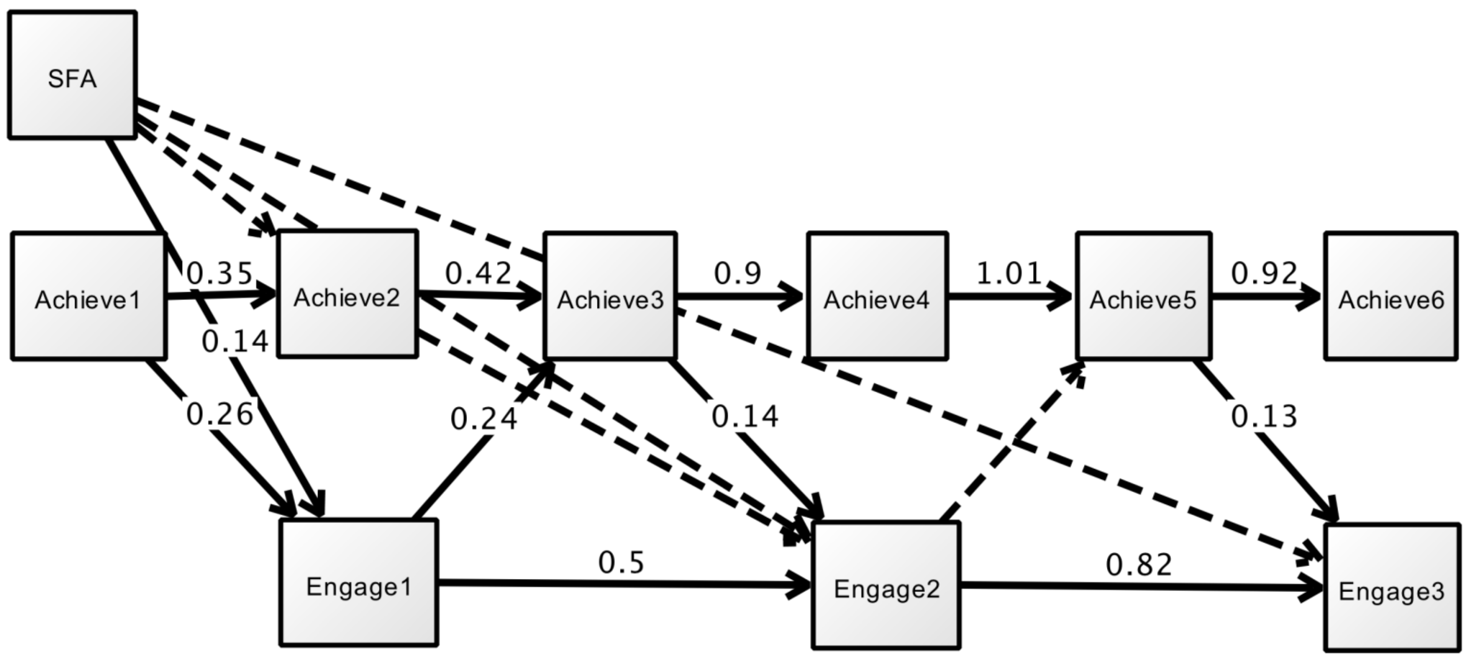

Figure C1. Completely standardized parameter estimates from Model 2. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, black, and propensity strata 3.

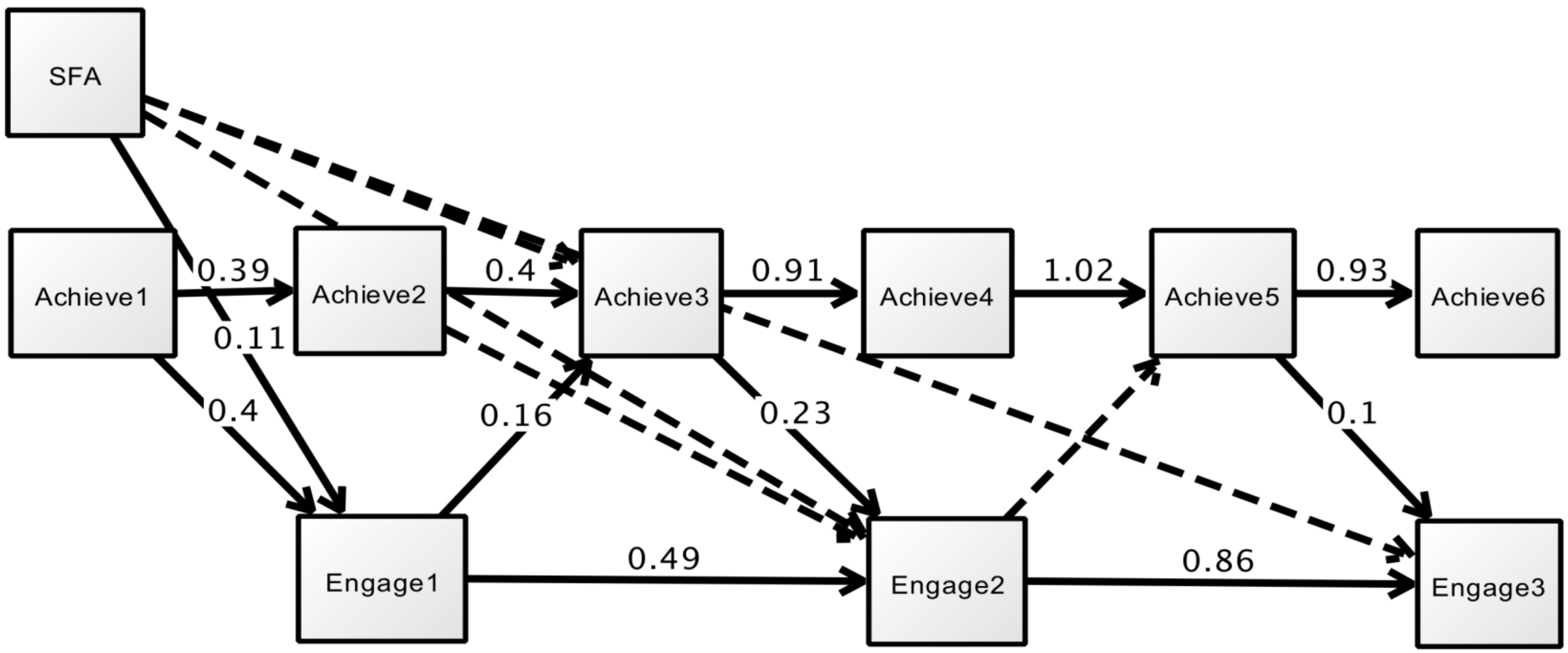

Figure C2. Completely standardized parameter estimates from Model 3. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, black, and propensity strata 3.

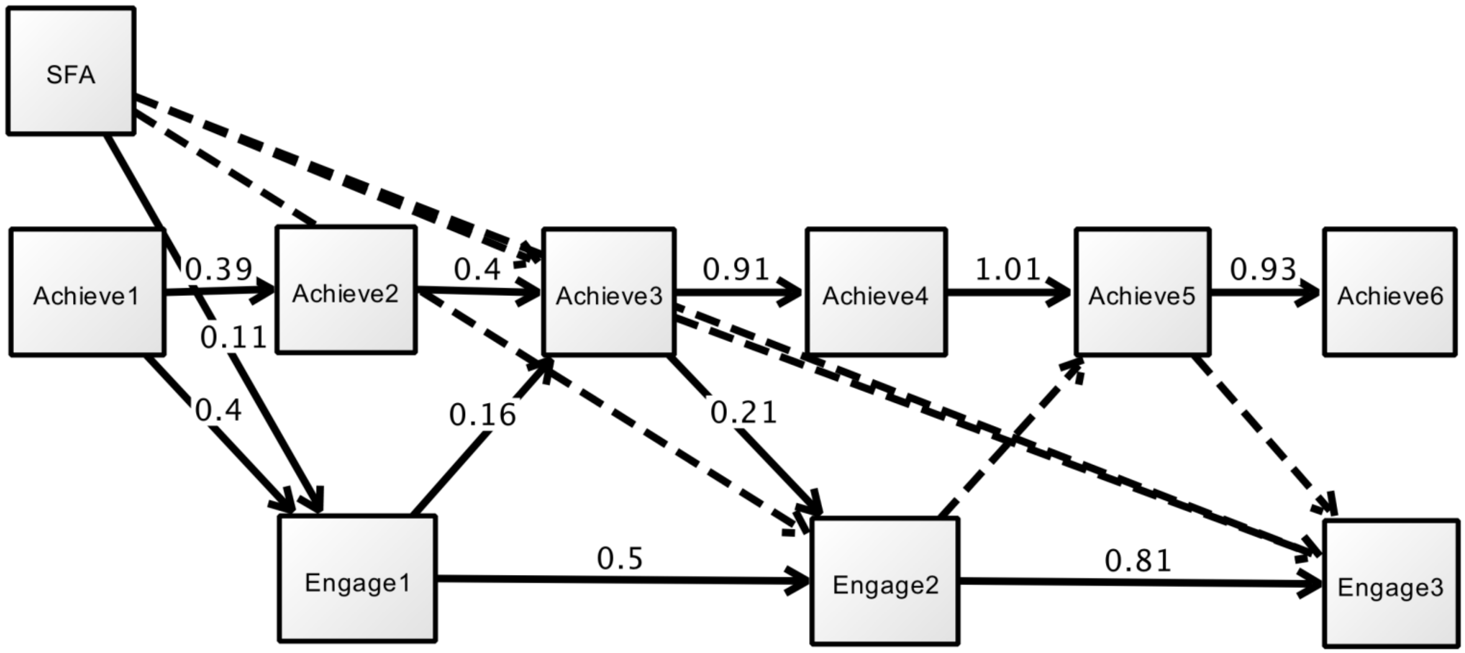

Figure C3. Completely standardized parameter estimates from Model 4. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, black, and propensity strata 3.

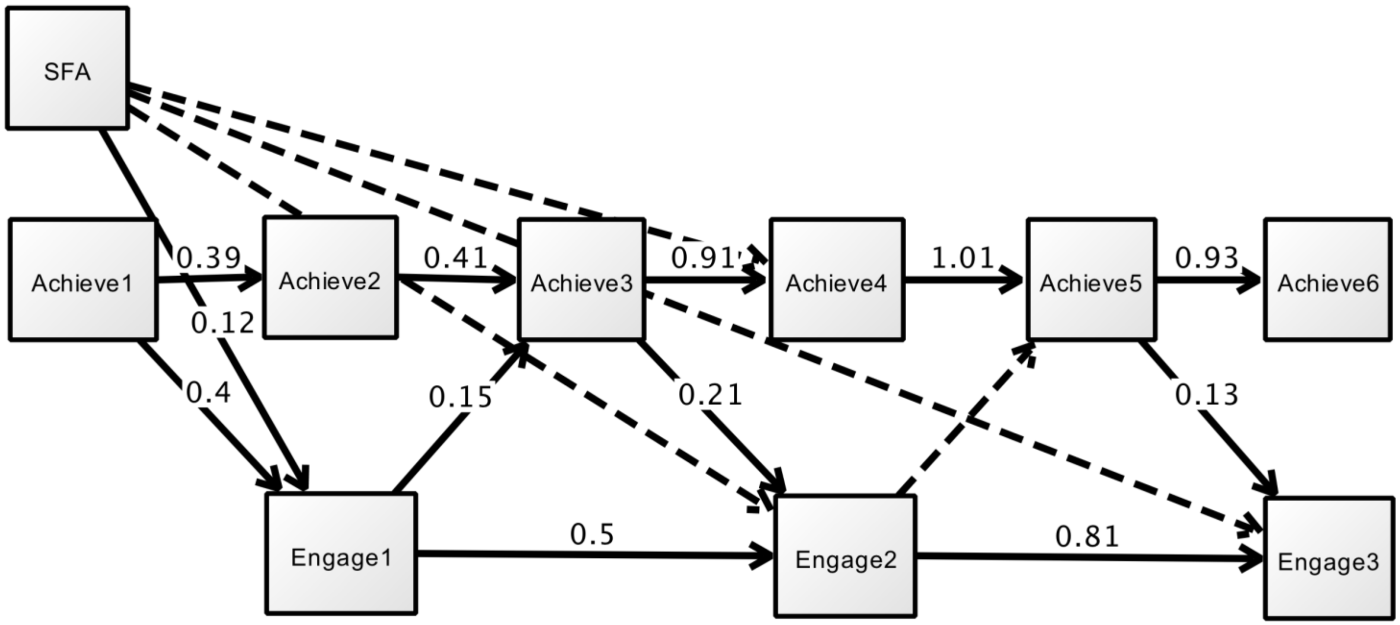

Figure C4. Completely standardized parameter estimates from Model 5. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, black, and propensity strata 3.

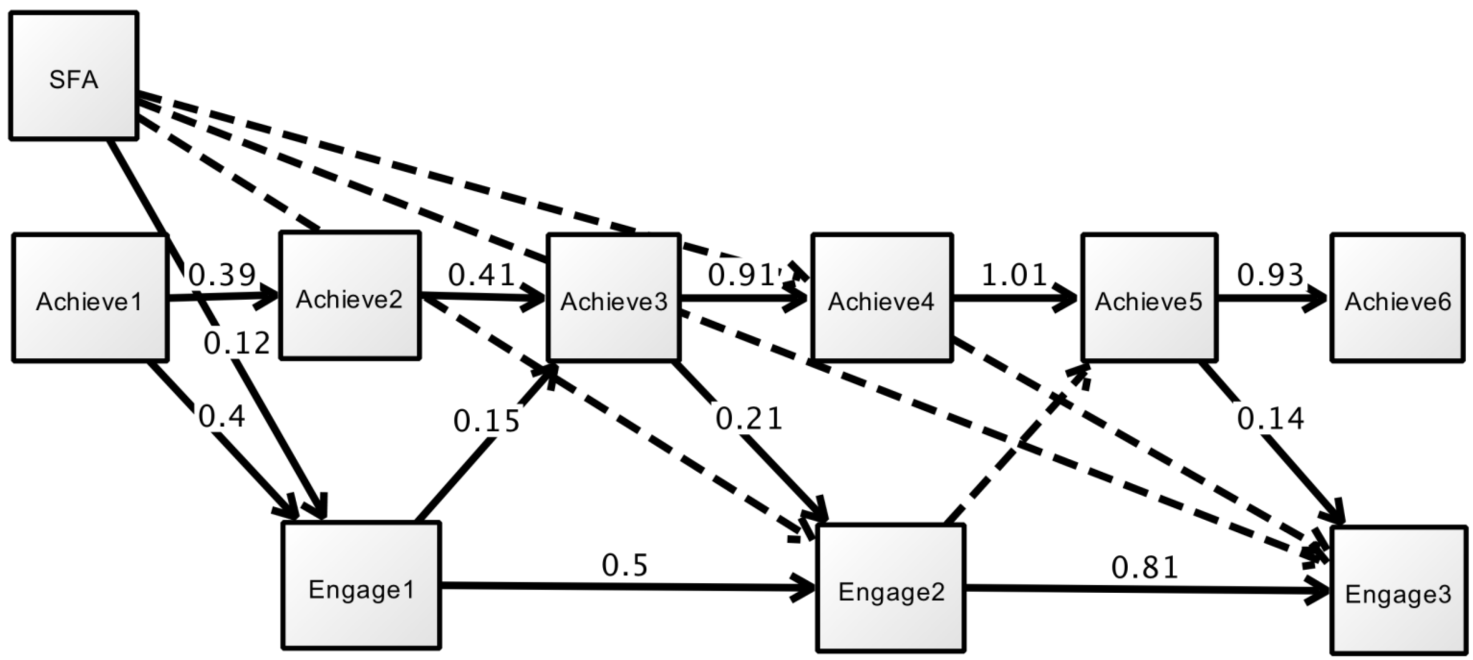

Figure C5. Completely standardized parameter estimates from Model 6. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships. Significant covariates included sex, SES, black, and propensity strata 3.

Appendix D

Table D1.

Standardized estimates for ARCL models for other non-cognitive outcomes

	Student-reported engagement		Teacher-reported antisocial behavior		Student-reported self-efficacy	
	β(SE)	95% CI	β(SE)	95% CI	β(SE)	95% CI
Achieve1→Achieve2	0.99**	[0.67,1.31]	0.38**	[0.27,0.49]	0.37**	[0.27,0.47]
Achieve2→Achieve3	1.29**	[1.03,1.55]	1.22**	[0.93,1.58]	1.28**	[1.04,1.51]
Achieve3→Achieve4	0.91**	[0.78,1.04]	0.89**	[0.69, 1.03]	0.89**	[0.76,1.01]
Achieve4→Achieve5	0.99**	[0.88,1.11]	1.05**	[0.89,1.23]	0.63**	[0.55, 0.71]
Achieve5→Achieve6	0.23**	[0.07,0.39]	0.64**	[0.48, 0.71]	0.32**	[0.18,0.46]
Noncog1→ Noncog 2	0.17**	[0.06, 0.29]	0.64**	[0.53,0.74]	0.16**	[0.05,0.27]
Noncog 2 → Noncog3	0.81**	[0.34,1.28]	0.86**	[0.67,1.04]	0.25**	[0.12,0.37]
Achieve1→ Noncog1	0.15**	[0.06,0.24]	0.06	[-0.02, 0.11]	0.15**	[0.10,0.26]
Achieve3→ Noncog2	0.25**	[0.17,0.33]	0.09**	[0.01, 0.11]	0.18**	[0.10,0.26]
Achieve5→ Noncog3	0.06	[-0.08,0.20]	0.09*	[-0.01,0.13]	0.12*	[-0.01,0.24]
Noncog1→Achieve3	-0.07	[-0.19,0.06]	0.09**	[0.01,0.25]	0.01	[-0.08,0.10]
Noncog 2→Achieve5	0.06	[-0.03, 0.14]	0.01	[-0.08,0.10]	0.12**	[0.06,0.19]
$\chi 2$/df	33.67/17**		104.87/20**		84.53/19**	
RMSEA [CI]	0.033		0.07[0.01,0.09]		0.06	
CFI	0.99		0.95		0.94	
TLI	0.97		0.90		0.89	
SRMR	0.04		0.06		0.06	

Note. *p<.05, **p<.01

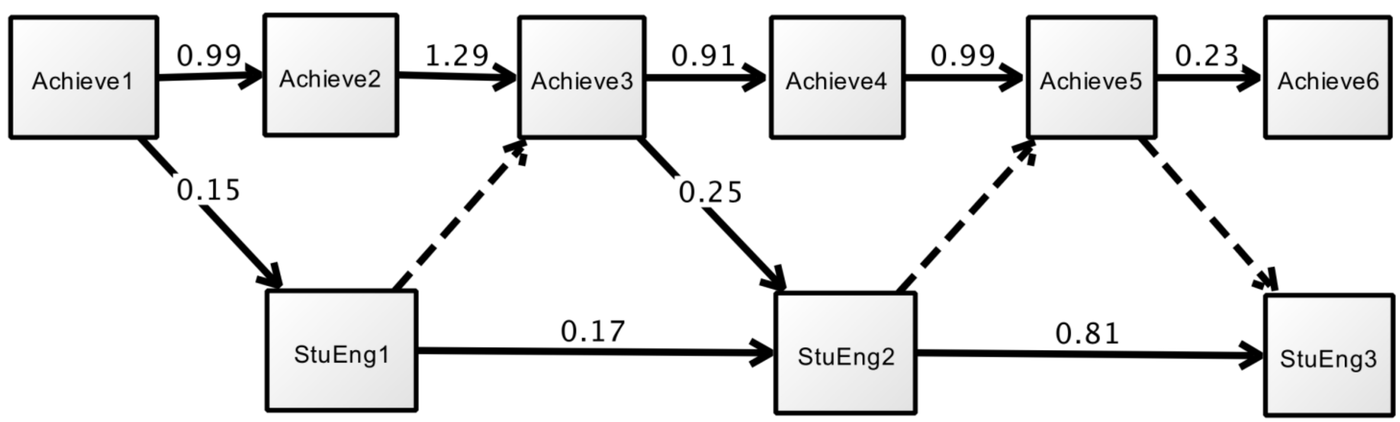

Figure D1. Completely standardized parameter estimates from ARCL model with student-reported engagement as outcome. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships.

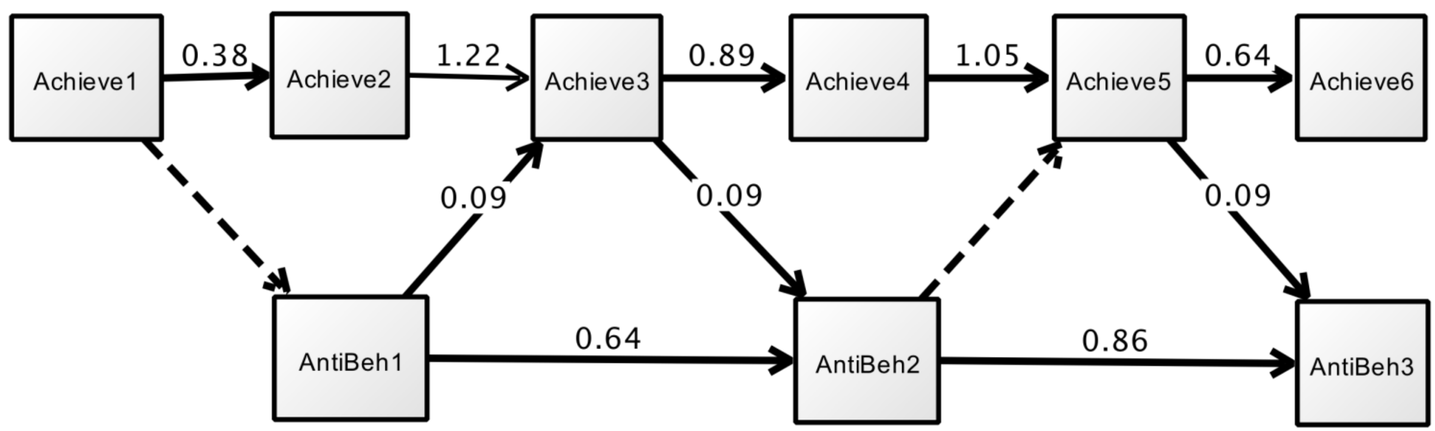

Figure D2. Completely standardized parameter estimates from ARCL model with teacher-reported anti-social behavior as outcome. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships.

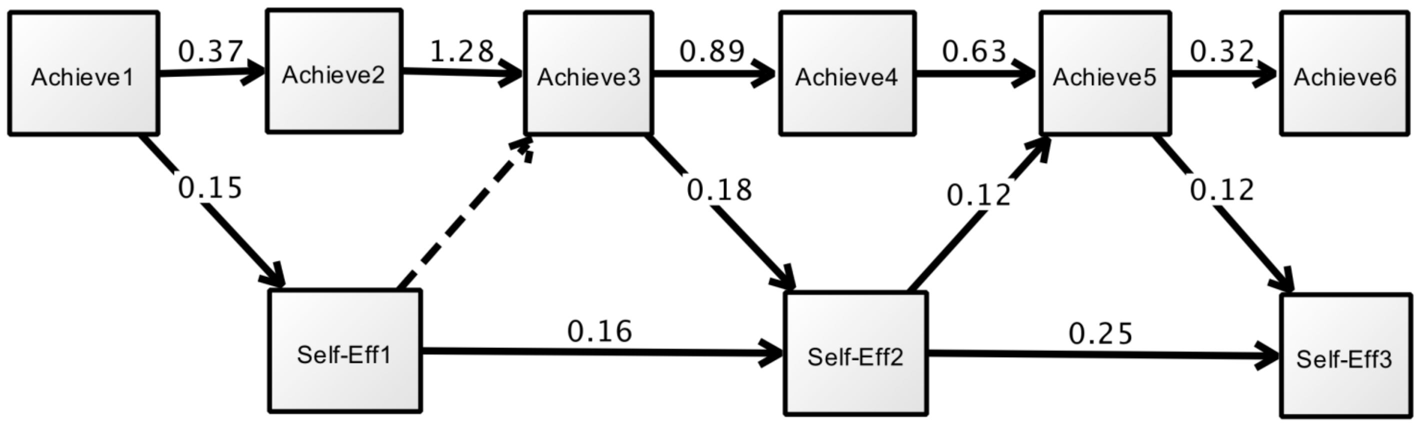

Figure D3. Completely standardized parameter estimates from ARCL model with student-reported self-efficacy as outcome. Covariate paths and covariances have been omitted for simplicity. A dashed line indicates non-significant relationships.

Appendix E

Curriculum Vitae/ Biography

ELIZABETH H. KIM

• 4112 QUEENSBURY RD, HYATTSVILLE MD 20781 • 510-292-9104 • lizkim@jhu.edu

EDUCATION/BIOGRAPHY

2013 - 2018	Ph.D, Johns Hopkins University School of Education
2011 - 2012	M.Ed in Learning and Teaching, Harvard Graduate School of Education
2008 – 2011	B.A.(magna cum laude) in English, Minor in Education, University of California, Berkeley
1990	Born- Orange County, California

PUBLICATIONS

Peer-reviewed journal articles:

Kim, E., & Corcoran, R. P. (2018). Factors that influence pre-service teachers' persistence. *Teaching and Teacher Education*, *70*, 204-214.

Davis, M. H., McPartland, J., Pryseski, C., & **Kim, E**. (in press). The effects of coaching on English teachers' reading instruction practices and students' reading comprehension. *Literacy Research and Instruction*.

Kim, E., & Corcoran, R. P. (2017). How engaged are pre-service teachers in the United States? *Teaching and Teacher Education*, *66*, 12-23.

Cheung, A., Slavin, R., **Kim, E.**, & Lake, C. (2017). Effective secondary science programs: A best evidence synthesis. *Journal of Research in Science Teaching, 54*(1), 58-81.

Corcoran, R. P., Cheung, A., **Kim, E**., & Xie, C. (2017). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*. https://doi.org/10.1016/j.edurev.2017.12.001

Peer-reviewed Conference Papers:

Kim, E., & Park, E. K., & Slavin, R. E. (2018). *Effective SEL Programs for Student Behavior*. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Kim, E., & Slavin, R. E. (2018). *The Role of Previous Achievement in Predicting Non-Cognitive Outcomes.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Kim, E., Davis, M., Gamez-Fusari, R., & Min, H. (2017). *Necessary but not sufficient: Content area literacy strategies in the Common Core era.* Paper presentation at the annual meeting of the American Educational Research Association, San Antonio, TX.

Corcoran, R.P. & **Kim, E.** (2017). *Student engagement among prospective teachers.* Presentation at the annual meeting of the American Psychological Association, Washington, DC.

Corcoran, R. P., Eisinger, J., **Kim**, E., & Ross, S. M. (2016). *Preparing students for the Common Core State Standards in reading: An evaluation of the McGraw-Hill Reading Wonders program.* Paper presentation at the annual meeting of the American Educational Research Association, Washington, DC.

Cheung, A., Slavin, R., Lake, C., **Kim, E**. (2016). *Effective secondary science programs: A best evidence synthesis*, Paper presentation at the annual meeting of the Society for Research on Educational Effectiveness. Washington, DC.

Research Reports and Other:

Slavin, R. E., & **Kim, E.** (2017). Reviewing social and emotional learning for ESSA: MOOSES, not parrots [Blog post]. Retrieved from https://www.huffingtonpost.com/entry/reviewing-social-and-emotional-learning-for-essa-mooses_us_591d9505e4b07617ae4cb9a0

Corcoran, R. P., Eisinger, J.M., & **Kim, E.** (2015). Evaluation of the McGraw-Hill Education Reading Wonder Program: Final Report. Retrieved from http://www.academia.edu/12349242/An_Evaluation_of_the_McGraw-Hill_Education_Reading_Wonders_Program_Interim_Report

Cheung, A., Slavin, R., **Kim, E.** & Lake, C. (2015). Effective Secondary Science Programs: A Best Evidence Synthesis. Retrieved from: http://www.bestevidence.org/word/sec_science_June_11_2015.pdf

RESEARCH EXPERIENCE

2015 - 2016 Research assistant, "Effective Programs for Social and Emotional Learning: A Systematic Review," Principal Investigator: Corcoran, R.P. Co-Investigator: Slavin, R.E. Sponsored by the Jacobs Foundation, $49,500, Funded - Active.

2014 - 2015 Research assistant, "Equipping High School Teachers to Increase Student Motivation and Course Passing Rates," Principal Investigator: Mac Iver, M. Co-Investigator: Mac Iver, D.J. Sponsored by Institute of Education Sciences, $1,500,000, Funded – Active.

Research assistant, "Improving Teaching and Learning: Development and Evaluation of the Blended Coaching Early Emotional Competencies Program Approach in Grades 3– 4," Principal Investigator: Corcoran, R.P. Sponsored by Character Lab, $300,000, Pending.

Research assistant, "Best Evidence Encyclopedia," Principal Investigator: Slavin, R.E. Sponsored by Institute of Education Sciences, US Department of Education, $9,997,674, Funded – Inactive.

Research assistant, "Randomized Controlled Trial Study of the McGraw-Hill Reading Wonders Program." Principal Investigator: Corcoran, R.P. Sponsored by McGraw Hill Education, Private Profit, $180,000, Funded – Inactive.

2013 - 2014 Research assistant, "Getting Students to the Finish Line: An Efficacy Study of a Ninth Grade Early Warning Indicator Intervention," Principal Investigator: Balfanz, R. Co-Investigator: Davis, M. Sponsored by Institute of Education Science, US Department of Education, $3,458,989, Funded – Active.

Research assistant, Self-Regulated Strategy Development Writing Project, Johns Hopkins University. Principal Investigator: Sandmel, K.

2011 - 2012 Research assistant, Good Work Project, Harvard Project Zero. Principal Investigator: Gardner, H.

TEACHING EXPERIENCE

Fall 2015 Teaching assistant, Research Methods and Systematic Inquiry II: Mixed Methods Research (Ed.D course)

Spring 2015 Teaching assistant, Evaluation of Education Policies and Programs (Ed.D course)

STATISTICAL/METHODOLOGICAL TRAINING

Relevant coursework:

- Controversies in Measurement for Education Science Research (Instructor: Dr. Julia Burdick-Will)
- Hierarchical Linear Models Independent Study (Instructor: Dr. Julia Burdick-Will)
- Hierarchical Linear Models (Instructor: Dr. Lieny Jeon)
- Mixed Methods Research (Instructor: Dr. Marcia Davis)
- Causal Inference (Instructor: Dr. Steve Morgan)
- Applied Multiple Regression Analysis (Dr. Amy Shelton)
- Qualitative Research Methods (Instructor: Dr. Carolyn Parker)
- Quantitative Research Methods (Instructor: Dr. Marc Stein)
- Basic and Inferential Statistics (Instructor: Dr. Deborah Carran)

PROFESSIONAL MEMBERSHIP

Member of AERA Division C: Teaching and Learning
Member of AERA Division K: Teaching and Teacher Education
Member of Phi Beta Kappa Honor Society (University of California, Berkeley)